IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants	:	Barany et al.	)	Examiner:
			)	S. Liu
Serial No.	:	09/963,698	)	
			)	Art Unit:
Cnfrm. No.	:	2018	)	1639
			)	
Filed	:	September 26, 2001	)	
			)	
For	:	DETECTION OF NUCLEIC ACID	)	
		SEQUENCE DIFFERENCES USING THE	)	
		LIGASE DETECTION REACTION WITH	)	
		ADDRESSABLE ARRAYS	)	
			)	

DECLARATION OF FRANCIS BARANY, Ph.D.
UNDER 37 C.F.R. § 1.132

Mail Stop RCE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

I, FRANCIS BARANY, Ph.D., hereby declare:

1.   I am a co-inventor of the above-identified application.

2.   I am currently a Professor in the Department of Microbiology and Immunology, as well as the program of Biochemistry and Structural Biology at Weill Medical College of Cornell University in New York, New York. I was also concurrently Adjunct Professor at The Rockefeller University, New York, New York, as well as Director of Mutation Research at the Strang Cancer Prevention Center at Weill Medical College of Cornell.

3.   I received a B.A. in Chemistry from the University of Illinois at Chicago, Chicago, Illinois in 1976 and a Ph.D. in Microbiology from The Rockefeller University, New York, New York in 1981. I conducted postdoctoral work from 1981-1982 in microbiology at The Rockefeller University and from 1982 to 1985 in molecular biology at John Hopkins University School of Medicine in Baltimore, Maryland.

4. My laboratory collaborates extensively with both academic and industrial researchers, including Memorial Sloan Kettering, Rockefeller University, UMDNJ, Princeton University, Weizmann Institute, Louisiana State University, Centers for Disease Control, and others. Our specialty is using thermostable enzymes and advanced DNA technology to help detect and characterize molecular changes in both cancers and infectious diseases. My laboratory has developed sensitive and specific assays using thermostable ligase combined with polymerase chain reaction ("PCR") for these applications. My laboratory is best known for developing the ligase chain reaction ("LCR"), ligase detection reaction ("LDR") and a programmable DNA chip (Universal Array). Other advances include the EndoV/Ligase mutation scanning assay and harmonized p53 mutation detection. I have served as the chair of the following National Institutes of Health Review Panels: Partnerships for Point of Care (POC) Diagnostic Technologies, Partnerships for Biodefense Food- and Water-borne Diseases, and Innovative Technologies for the Molecular Analysis of Cancer. I am inventor on over two dozen issued patents, comprising more than 10% of issued patents at Weill Cornell Medical College, and have authored over 100 peer-reviewed articles. In 2004, I was honored as Medical Diagnostics Research leader, Scientific American 50. A copy of my Curriculum Vitae, listing these patents and publications, is attached as Exhibit 1 hereto

5. Prior to the filing date of my above referenced patent application, it was well appreciated in the art that the ability to accurately identify low abundance nucleic acid sequence variations, including single nucleotide polymorphisms, insertions, deletions, or translocations at multiple adjacent, nearby, and distant genomic loci would have profound implications for the identification of genetic disorders, the diagnosis and treatment of cancer, and the detection of infectious diseases. Cancer, for example, can arise from the accumulation of mutations in genes controlling cell cycle, apoptosis, and genome integrity. Oncogenes may be activated by point mutations, translocations, or gene amplification, while tumor suppressor genes may be inactivated by point mutations, frameshift mutation and deletions. These mutations may be inherited or somatic, arising from exposure to environmental factors or from malfunctions in DNA replication and repair machinery. Since the capacity to detect these cancer related mutations would significantly enhance cancer detection and diagnosis, and identify the most effective and targeted cancer treatment protocols, there was a well recognized need in the art for an assay that could achieve early and accurate detection of these cancer related mutations. I am presenting this declaration to

12922360.2

demonstrate how the efforts of others in the art to develop a hybridization array-based detection assay to meet this need have failed and continue to fail. I am also presenting this declaration to show how my array with capture oligonucleotides having greater than sixteen nucleotides and sequences selected to hybridize with complementary oligonucleotide target sequences under uniform hybridization conditions across the array of oligonucleotides with minimal cross reactivity, where each capture oligonucleotide of the array differs in sequence from other adjacent capture oligonucleotides, when aligned to each other, by at least 25% of the nucleotides (hereafter identified as "My Array Design"), has overcome these failures and has successfully resolved this unmet and long-felt need.

6.   The development of an assay suitable for the detection of cancer related mutations was fraught with challenges. The first challenge was to identify an approach that could detect very low abundant target mutations within a patient sample containing a plurality of closely related non-target sequences (i.e., normal, non-mutant sequence). In primary tumors for example, normal stromal cell contamination can be as high as 70% of total cells. Therefore, a mutation present in only one of the two chromosomes of a tumor cell may represent as little as 15% of the DNA sequence in a sample. In addition, early detection of such mutations requires the ability to detect as few as one mutant copy of a nucleic acid sequence in the presence of over 100 non-mutant copies of the nucleic acid sequence. Accordingly, the detection assay had to be highly sensitive. A second challenge was to develop a highly specific assay having the capacity to reproducibly discriminate and detect a plurality of often closely spaced mutations in multiple genes without generating false-positive or false-negative results. Finally, it was highly desirable to employ an assay that could achieve this highly sensitive and specific mutation detection in non-invasively collected patient samples to help reduce overall cost and, more importantly, alleviate patient discomfort.

7.   Although the advent of DNA array technology, based on direct hybridization of target sequence to the array, resulted in a paradigm shift in identifying expression changes and known SNPs on a genomic scale, it failed, and, as discussed below, continues to fail to meet the above noted challenges in detecting mutations.

8.	Typical DNA hybridization arrays are designed to simultaneously discriminate and detect multiple target sequences differing in sequence by only one or a few nucleotides. Target sequence discrimination using a hybridization array depends on the highly specific binding affinity of the immobilized capture oligonucleotides to their complementary labeled target sequences. The hybridized labeled target sequences are subsequently detected and identified by their location of hybridization on the array surface.
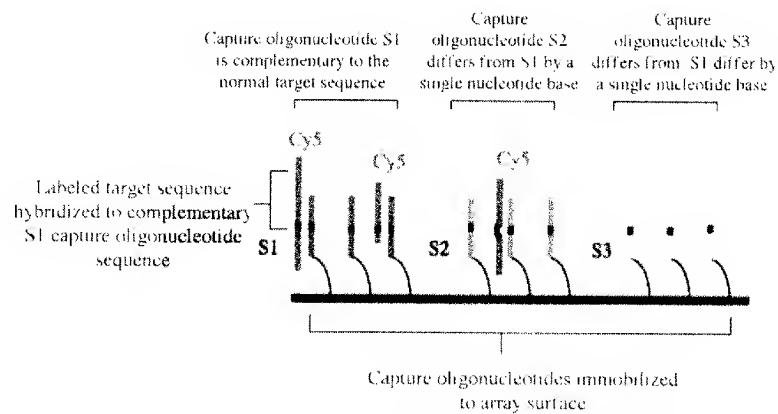
Figure 1: A typical direct hybridization array having capture oligonucleotides designed to simultaneously discriminate and detect multiple nucleotide variations at multiple adjacent and nearby genetic loci.

9.	Figure 1 depicts a typical hybridization array having capture oligonucleotide probes immobilized on the array surface where the capture probes are designed to discriminate nucleotide variations (*i.e.*, allelic variations) at multiple adjacent and nearby loci (*e.g.*, S1, S2, and S3). The capture oligonucleotides responsible for allelic discrimination at a particular locus (*e.g.*, S1, S2, or S3 in Figure 1) differ from each other by only a single nucleotide base substitution, insertion, or deletion, and, therefore, have very similar nucleotide sequences as shown in Figure 2 below. Consequently, these capture probes also have very similar melting temperatures (Tm).

```
Target: 5'-CATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGA          Probes:
Probes:         3'-TTTATAGTAGAAACC (SEQ. ID NO:9G)              93%
                                                            Identical
                3'-TTTATAATAGAAACC (SEQ. ID NO:9A)

                3'-TTTATATTAGAAACC (SEQ. ID NO:9T)

                3'-TTTATACTAGAAACC (SEQ. ID NO:9C)
                                                              87%
Target: 5'-CATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGA         Identical
               3'-TTATAGTAGAAACCA (SEQ. ID NO:10T)

               3'-TTATAGCAGAAACCA (SEQ. ID NO:10C)
Target: 5'-CATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGA
               3'-TTATAGGAGAAACCA (SEQ. ID NO:10G)            93%
                                                            Identical
               3'-TTATAGAAGAAACCA (SEQ. ID NO:10A)
```

Figure 2: Capture oligonucleotide probes of a typical hybridization array. Figure 2A shows that capture oligonucleotide probes designed to detect target sequence variations at one nucleotide position share 93% sequence identity while probes designed to detect target sequence variations at nearby genetic loci (i.e., probes of 2A and 2B) share 87% sequence similarity.

10. In addition to capture probes designed to detect sequence variations at a single locus or nearby loci, typical hybridization arrays also contain capture probes that are designed to detect distal mutations. Because capture probes are target-sequence specific, the nucleotide sequences of probes detecting distal target sequences will differ significantly in both sequence and melting temperature from other capture probes on the array designed to detect other distant mutations.

11. Figure 3 below illustrates some of the problematic results that emanate from the typical direct hybridization array designed to detect target sequences having overlapping sequence homology (e.g., target sequences that have only single nucleotide differences). It is possible for a single target to bind to multiple oligonucleotide probes with different, yet similar sequences due to mismatched cross hybridization. Cross hybridization will result in the generation of false-positive signals.
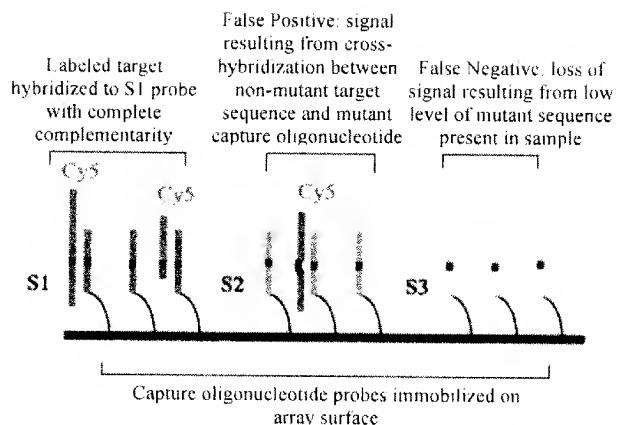
Figure 3. Problematic results of a typical hybridization array

12. Another major drawback of hybridization arrays is their propensity to generate false-negative signals. As noted above, significant variability in nucleotide sequence and melting temperature exists between capture probes that are designed to detect distant mutations. Since optimal hybridization conditions for a target and its complementary capture probe are sequence specific, employing uniform, highly stringent hybridization conditions across the array that are suitable for all capture probe-target sequence pairs is difficult, if not impossible. The application of hybridization conditions that are overly stringent for some sequences will prevent target-probe hybridization and lead to the loss of signal (*i.e.*, false-negative signal). Since the melting temperature between target and probe sequence varies across the array, stringent hybridization conditions will result in weak or missing signal from low abundant mutations. If non-stringent conditions are used to detect low level mutations, this will significantly increase the likelihood of cross-hybridization leading to false-positives.

13. In addition to the possibility that a single target will bind to multiple oligonucleotide probes with different, yet similar sequences due to mismatched cross hybridization, there are many other competitive processes that influence signal intensity values generated during array hybridization. Besides the desired target binding to probe (Figure 4A), there is the undesired probe self binding (Figure 4B), folding of target to reduce binding to probe (Figure 4C), and dimerization of adjacent probes (Figure 4D) as illustrated below.
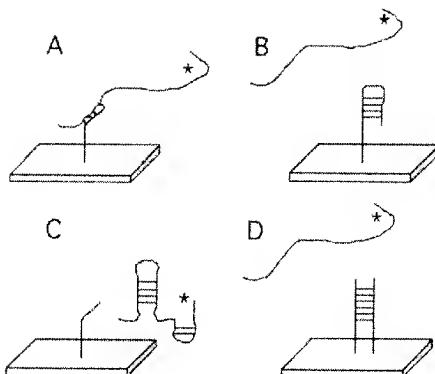
Figure 4: Depiction of four competitive processes on signal intensity values. Each panel shows a labeled (*) target and an immobilized probe on a microarray. (**A**) hybridization of a target to a probe; (**B**) probe self-folding; (**C**) folding of the target and (**D**) dimerization of adjacent probes (reproduced from Pozhitkov et al., *Nucleic Acids Research* 34(9):e66 (2006))

14.   While a great deal of effort has been invested in developing design strategies that generate capture probes having minimal cross-reactivity to non-complementary target sequence, minimal cross-reactivity to other probe sequences, and do not undergo self-folding, none of these efforts have proven successful. These design strategies, typically based on the thermodynamic properties of the capture oligonucleotides (*e.g.*, guanine-cytosine content, secondary structure, melting temperature, etc.), attempted to predict oligonucleotide duplex formation. Using these strategies, capture probes were designed so that a single mismatch base pair would, theoretically, significantly lower the binding affinity of the mismatched duplex compared to the corresponding perfectly matched duplex at a given temperature. These differential binding affinities of a target sequence to a mismatch or perfect match probe provide the basis of sequence discrimination, allowing for the identification of target sequence because it is bound to a perfect match but not a mismatched probe at a specified temperature.

15.   Figure 5 below depicts the melting curve profiles, *i.e.*, the temperature-dependent dissociation of capture probe bound to its target sequence, for a set of mismatch and perfect match probes. Ideally, the melting profiles of the mismatch probe and perfect match probe would be tight to facilitate unambiguous differential signal detection. However, as indicated by the red arrow in the top panel, some mismatched probes have melting curves that overlap exactly with the perfect matched probes, making it impossible to distinguish target binding to the mismatched probe from the perfect match probes. The first derivatives of these melting curves shown in the bottom panel of Figure 5, more clearly

illustrates the overlap in melting temperatures between the mismatch and perfect match probes (red arrow). In addition, it shows that some mismatched probes produce signal over a broad temperature range, generating signal that overlaps with both mismatch and perfect match probes (*see* bottom panel of Figure 5). These mismatch probes also cannot be accurately distinguished from perfect match probes.
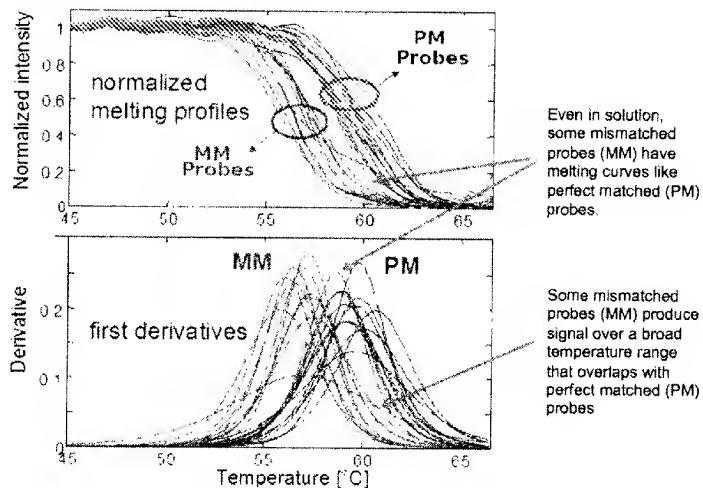

Figure 5: Melting profiles of mismatch (MM) and perfect match (PM) oligonucleotides probes in solution (reproduced from www.genewave.com/images/manager/hyblive_schemal.jpg). The top panel is a plot monitoring the decrease in signal intensity that occurs as the DNA duplex melts with increasing temperature. The bottom panel is a plot of the negative first derivative of the change in fluorescence (-dF/dT, the rate of change of fluorescence) versus temperature. The distinct peaks in this plot correspond to the melting temperature of each DNA duplex.

16.   Naiser et al., "Impact of Point-Mutations on the Hybridization Affinity of Surface-Bound DNA/DNA and RNA/DNA Oligonucleotide-Duplexes: Comparison of Single Base Mismatches and Base Bulges," *BMC Biotech.* 8:48 (2008) ("Naiser") (attached hereto as Exhibit 2) describes a comprehensive analysis of how single nucleotide variations (referred to as "point defects") affect the hybridization of fluorescently labeled oligonucleotide targets to surface-bound oligonucleotide probes (Naiser at p. 49, col. 2, para. 2). Naiser generated a set of point mutated probes derived from a common probe sequence motif that was complementary to a region of a target sequence (*id.* at figure legend of Figure 1). Exemplary probe sequences representing single nucleotide substitutions, insertions, and deletions at the first two bases (*i.e.*, first two defect positions) of the probe sequence are depicted in Figure 1 of Naiser (reproduced below as Figure 6).
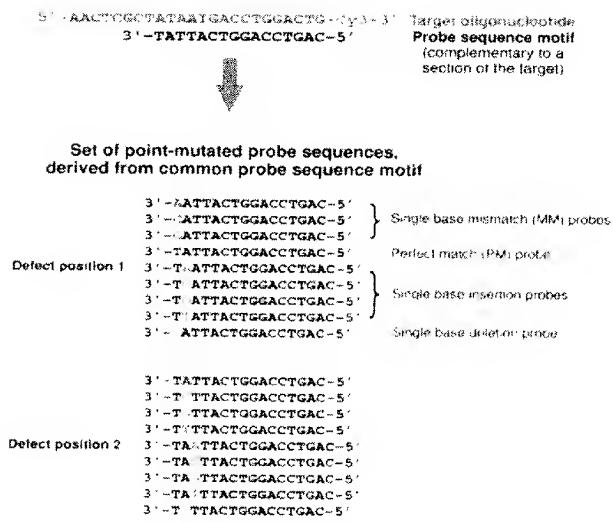
Figure 6: Figure 1 of Naiser et al., *BMC Biotech.* 8:48 (2008) showing the comprehensive set of point-mutated probes is derived from a common probe sequence motif which is complementary to the target sequence. Probe sequences are shown for the first two defect positions only.

17.   The probe sequences were arranged on a microarray surface as a compact feature block and a sample containing a single target nucleotide sequence was contacted with the array surface to facilitate target-probe hybridization (*id.* at p. 49, col. 2, para. 2, and Figure 1). Hybridization signals resulting from target sequence hybridization to individual probes in the probe set were plotted against the position of the defect in the probe sequence to create a defect profile (*id.* at figure legend of Figure 1). The defect profile shown in Figure 7 (below) provides a direct comparison of the binding affinities for a plurality of mismatch oligonucleotide duplexes (*i.e.*, duplexes between target and probe where the probes differ from the target sequence by a single base mismatch, insertion, or deletion), and demonstrate the considerable variability in binding affinities that exist between mismatched probe sequences.

19. Figure 8 below is a fluorescence micrograph of the microarray feature-block illustrating the problem associated with trying to accurately discriminate mismatch and perfect match probe binding to target sequence based on hybridization signal intensity. The microarray feature-block comprises variations of a 16mer probe sequence motif (*i.e.*, probe sequences varying by single base insertions, deletions, and substitutions) subject to hybridization to a single nucleic acid target sequence. Each 3 x 3 sub-array comprises one perfect matching probe, three single base mismatch probes, four insertion probes, and one single base deletion probe. While the signal intensity generated by the perfect match probe is distinguishable from the signal intensity generated by the mismatch probes in the center two sub-arrays outlined in red (*i.e.*, the single brightest square within each sub-array represents the intensity of the perfect match probe), there is still considerable signal observed at other positions. Further, where mismatches are elsewhere, it is even more difficult to distinguish signal intensities generated by perfect match and mismatch probes.
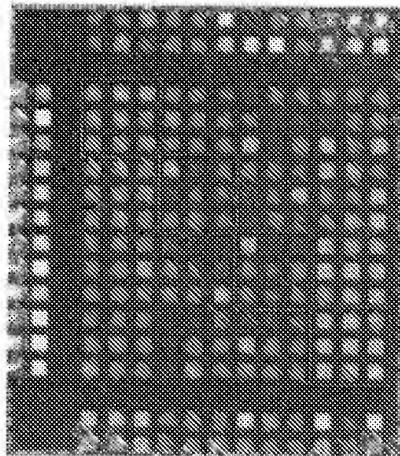
Figure 8: Fluorescence micrograph of a microarray feature-block comprising variations of the 16 mer probe sequence motif 3'-TATTACTGGACCTGAC-5'. Microarray hybridization was performed with the 5'-Cy3-labeled RNA oligonucleotide target 3'-AACUCGCUAUAAUGACCUGGACUG-5' (target concentration: 1 nM in 5 × SSPE, pH 7.4, 0.01% Tween-20, T = 30°C). Each 3 × 3 sub-array comprises one perfect matching probe, three single base mismatch probes, four insertion probes and one single base deletion probe. Figure reproduced from Naiser et al., "Position Dependent Mismatch Discrimination of DNA Microarrays – Experiments and Model," *BMC Bioinformatics* 9:509 (2008) (attached hereto as Exhibit 3).

20. Naiser's findings are consistent with those reported in an earlier study by Pozhitkov et al., "Test of rRNA Hybridization to Microarrays Suggest that Hybridization Characteristics of Oligonucleotide Probes for Species Discrimination Cannot be Predicted," *Nucleic Acids Research* 34(9):e66 (2006) ("Pozhitkov") (attached hereto as Exhibit 4).

Pozhitkov assessed the utility of *in silico* predictions of probe-target duplex stabilities using DNA arrays for the detection of rRNA sequences (Pozhitkov at p. 2, col. 1, para. 3).

21.   Pozhitkov's assessment of the effects of mismatches in the probe sequence on signal intensity values generated when hybridized to the non-mismatched target sequence also revealed that mismatch position, mismatch type, and the type of neighboring nucleotides surrounding the defect have significant effects on the normalized signal intensity values *(id.* at p. 7, col. 1, last para.). "Moving the MM base away from the 5' or 3' termini to the center of the probe significantly decreases signal intensities (Figure 3). ... However, we emphasize that this was an average result, and note that in some individual cases, MM probes with central mismatches (positions 9–11) were observed to have signal intensities that were *equal to or 1.6 times higher* than the corresponding perfectly matched probe" *(id.* at p. 7, col. 2., para. 2).

22.   The implications of Pozhitkov's and Naiser's findings are that direct hybridization methods which attempt to simultaneously discriminate and detect nucleic acid sequence variations are inadequate because of the unpredictable cross-hybridization between target sequence and mismatch or perfect match probe sequences. Thermodynamic parameters are simply not capable of accurately predicting mismatch and perfect match oligonucleotide duplex formation. Accordingly, despite tremendous efforts to improve the reliability and reproducibility of the technology, the findings of Naiser and Pozhitkov described above, clearly indicate that a fundamental understanding of the technology is lacking and that even current approaches for microarray design are inadequate. Likewise, this same problem extends to any assay format where the target sequence is detected and distinguished from other sequences by its hybridization to a complementary sequence.

Summary of U.S. Patent No. 5,510,270 to Fodor et al. ("Fodor")

23.   Fodor relates to a method for synthesizing and screening polymers on a solid substrate (Fodor at abstract). The method involves providing a substrate which may include linker molecules on its surface *(id.* at col. 8, lines 46–48). On the substrate or a distal end of the linker molecules, a functional group with a protective group is provided *(id.* at lines 58–59). The protective group may be removed upon exposure to radiation, electric fields, electric currents, or other activators to expose the functional group *(id.* at lines 59–62).

Concurrently or after exposure of a known region of the substrate to light, the surface is contacted with a first monomer unit M1 which reacts with the functional group which has been exposed by the deprotection step (*id.* at col. 9, lines 11–14). Thereafter, second regions of the surface (which may include the first region) are exposed to light and contacted with a second monomer M2 (which may or may not be the same as M1) having a protective group (*id.* at lines 26–29). These steps are repeated until the substrate includes desired polymers of desired lengths (*id.* at lines 37–38). Monomers may include amino acids, nucleotides, pentoses, and hexoses (*id.* at col. 6, lines 14–18).

24.   Fodor does not teach arrays of oligonucleotides on a solid support where each oligonucleotide of the array differs in sequence from other adjacent oligonucleotides when aligned to each other by at least 25% of the nucleotides.

Summary of U.S. Patent No. 5,474,796 to Brennan et al. ("Brennan")

25.   Brennan relates to an apparatus and methods for making arrays having functionalized binding sites on a support surface and conducting a large number of chemical reactions on the support surface (*see* Brennan at abstract and col. 2, lines 11–12). Brennan further relates to a method of determining or confirming the nucleotide sequence of a target nucleic acid where the target nucleic acid is labeled and hybridized to oligonucleotides of known sequence bound to sites on the array plate (*id.* at col. 3 lines 11–15).

26.   It is my understanding that the United States Patent and Trademark Office ("PTO") considers Brennan's disclosure of arrays having 3-mers and 10-mers attached thereto where every possible permutation of the 3-mer or 10-mer is provided, to be the same as each capture oligonucleotide of an array differing in sequence from other adjacent capture oligonucleotides when aligned to each other by at least 25%. I respectfully disagree for the reasons set forth below.

27.   Although Brennan teaches that the resulting 10-mer oligonucleotides on an array represent all permutations of the 10-mer sequence, each 10-mer oligonucleotide of the array does not differ in sequence from other adjacent 10-mer oligonucleotides, when aligned to each other, by at least 25% of the nucleotides. In fact, following the method of oligonucleotide synthesis taught by Brennan, each "oligonucleotide element, moving in a 5'-3' direction, is

12922360.2

identical to the preceding element in nucleotide sequence except that it deletes the 5'-most nucleotide and adds a 3'-most oligonucleotide" (Brennan at col. 9, lines 49–53). Therefore, adjacent oligonucleotides formed according to the method of Brennan have significant sequence similarity when aligned. In other words, nine out of ten nucleotides of adjacent 10mers are the same, so that adjacent 10mers have 90% sequence identity when aligned, differing by only 10% (see ClustalW2 pairwise sequence alignment results attached hereto as Exhibit 5).

Summary of U.S. Patent No. 5,594,121 to Froehler et al. ("Froehler")

28.   Froehler discloses oligomers containing 7-deaza-7-substituted purines and related analogs that have enhanced ability for double-and triple-helix formation with single- or double-stranded target nucleic acid sequences (Froehler at abstract). Such oligomer analog compositions can be used for diagnostic assays that employ methods where the oligomer or nucleic acid to be detected is covalently attached to a solid support (*id.* at col. 33, lines 18–21).

29.   Froehler teaches that oligomers (*e.g.*, dimers – hexamers) are useful as synthons (*i.e.*, structural unit within a unit) for producing longer oligomers (*id.* at col. 6, lines 64–65 and col. 7, lines 59–60). However, Froehler fails to teach oligomers on a solid support where each oligomer differs in sequence from other adjacent oligomers, when aligned to each other by at least 25% of the nucleotides.

30.   The combination of Fodor, Brennan, and Froehler does not teach arrays of oligonucleotides on a solid support where each capture oligonucleotide differs in sequence from other adjacent capture oligonucleotides, when aligned to each other by at least 25% and hybridize to complementary oligonucleotide target sequences under *uniform* hybridization conditions across the array of oligonucleotides.

Summary of U.S. Patent No. 5,527,681 to Holmes ("Holmes")

31.   Holmes relates to methods, devices, and compositions for synthesis and use of diverse molecular sequences on a substrate (Holmes at col. 1, lines 65–66). In particular, Holmes discloses the synthesis of an array of polymers in which individual

monomers in a lead polymer are systematically substituted with monomers from one or more basic sets of monomers (*id.* at col. 2, lines 1–4). On the substrate or a distal end of linker molecules, a functional group with a protective group is provided (*id.* at col. 7, lines 51–53). The protective group may be removed upon exposure to a chemical reagent, radiation, electric fields, electric currents, or other activators to expose the functional group (*id.* at lines 53–57). Concurrently or after exposure of a known region of the substrate to light, the surface is contacted with a first monomer unit M1 which reacts with the functional group which has been exposed by the deprotection step (*id.* at col. 8, lines 5–9). Thereafter, second regions of the surface (which may include the first region) are exposed to light and contacted with a second monomer M2 (which may or may not be the same as M1) having a protective group (*id.* at lines 21–26). These steps are repeated until the substrate includes desired polymers of desired lengths (*id.* at lines 38–39). Monomers may include amino acids, nucleotides, pentoses, and hexoses (*id.* at col. 4, lines 6–11).

32. Holmes, like Brennan and Froehler, does not teach arrays of oligonucleotides on a solid support where each oligonucleotide of the array differs in sequence from other adjacent oligonucleotides, when aligned to each other by at least 25%. Further, the combination of Holmes, Brennan, and Froehler fails to teach a method that such oligonucleotides are attached to a solid support and hybridize to complementary oligonucleotide target sequences under *uniform* hybridization conditions across the array of oligonucleotides.

Summary of the Present Invention:

33. My Array Design avoids all of the aforementioned problems associated with typical hybridization arrays (*i.e.*, target-capture probe cross-hybridization and false-positive/negative signal generation). Identifying one or more target nucleotide sequences using My Array Design may employ a ligase detection reaction (LDR) followed by a high-throughput method of detection to decouple mutation discrimination from hybridization and detection. Hybridization is carried out using divergent probe sequences that are not homologous to the target sequence being detected or any other known genomic sequence. Although divergent in sequence, these probes are carefully designed to have very similar hybridization properties. This strategy significantly reduces cross-hybridization to enhance the specificity of target discrimination, while allowing for the use of uniform hybridization conditions across the array to facilitate a high-throughput assay format.

34.   A schematic representation of one embodiment of my invention using My Array Design to detect single base changes in a gene is provided in Figure 13 below.

Figure 13: (A) LDR is performed using a common probe for each genetic locus which contains a unique addressable array-specific portion (Zp1 or Zp2) and allele-specific probes, each containing a unique detectable reporter label (*e.g.*, Cy3 or Cy5). Ligation of the two probes occurs only when there is perfect complementarity at the junction. (B) The multiplexed ligation products are captured on an addressable array containing capture probes that are complementary to the addressable array-specific portion of the common LDR probe.

35.   As illustrated in Figure 13A, a plurality of oligonucleotide probe sets are used, where each probe set is characterized by (a) a first oligonucleotide probe having a target specific portion and an addressable array specific portion (Zp1 or Zp2) that is distinct from the target sequence and different for each gene locus that is interrogated, and (b) a second oligonucleotide probe having an allele-specific target portion and a unique detectable reporter label portion (*e.g.*, Cy5 or Cy3). In an LDR process, the oligonucleotide probes are complementary to only one strand of the target nucleic acid as shown above, resulting in the linear amplification of the target nucleotide sequence.

36.   When oligonucleotide probes of a probe set hybridize adjacent to one another on a target sequence, ligation occurs only if there is perfect complementarity at the ligation junction. The resulting ligation product contains (a) the addressable array specific portion, (b) the target-specific portions, and (c) the detectable reporter label. The addressable

array-specific portion of a ligation product is complementary to a capture oligonucleotide immobilized at a particular site or "address" on the solid support. As depicted in Figure 13B above, the addressable array-specific sequences together with a detectable label can discriminate between a plurality of different target sequences.

37.   The plurality of capture oligonucleotides immobilized on a solid support are designed to differ substantially from each other in their nucleotide sequence, yet all have the same or similar melting temperature. This design strategy drastically minimizes any chance of cross-hybridization leading to false-positive signals, while allowing for simultaneous capture of a plurality ligation products, by their addressable array sequence, under uniform hybridization conditions across the array. In accordance with My Array Design, we have designed 24-mer capture oligonucleotides that differed from each other by at least 6 bases or at least 25% when aligned to each other based on sequence similarities (*see* Figure 14 below), yet have the same or very similar melting temperatures (Tm).

```
Probes: Zip 12 (2-4-4-6-1-1)=24 mer

Target: 3'-TAGC CCAT CCAT TGGA ACGC ACGC - LDR PRODUCT 1        24/24 match
        5'-ATCG GGTA GGTA ACCT TGCG TGCG-3' SEQ ID NO: 7     Yes hybridization

                                                                    Probes:
                                                              \     25% or more
                                                               \    Different
                                                               /
        Zip 14 (4-4-6-6-3-1)=24 mer                            /
Target: 3'-TAGC CCAT CCAT TGGA ACGC ACGC - LDR PRODUCT 1
        5'-GGTA GGTA ACCT ACCT CAGC TGCG-3' SEQ ID NO: 8        12/24 match
                                                              No hybridization


Target: 3'-TAGC CCAT CCAT TGGA AC C ACGC - LDR PRODUCT 1
                                    G                          13/24 match
        5'-GGTA GGTA ACCT ACCT CAGC TGCG-3' SEQ ID NO: 8     No hybridization
```

Figure 14. Capture oligonucleotides of the present invention are designed to differ from each other by at least 25% of their nucleotide sequence when aligned. Using this design strategy cross-hybridization between non-complementary addressable array portions and capture oligonucleotides will not occur.

38.   As illustrated in Figure 14, cross-hybridization between an addressable array sequence and the wrong capture oligonucleotide probe sequence will not occur because of the extent of non-complementarity that exists between them. Because the capture oligonucleotide sequences remain constant (*i.e.*, the sequence is not target-specific), and their complements can be appended to any set of LDR primers, the addressable arrays of My Array Design have universal application.

39.   As summarized above, My Array Design provides for the highly sensitive and specific detection and discrimination of target sequences that differ by only a single nucleotide substitution, deletion, or insertion in a sample. As summarized below and described in detail in the attached peer-reviewed publications, My Array Design provides a rapid and reliable method for the detection of genomic mutations (*e.g.*, genetic disease mutations and cancer related mutations), promoter methylation, and infectious diseases *(e.g.*, bacterial, fungal and viral infections).

Cancer Detection

40.   Gerry et al., "Universal DNA Microarray Method for Multiplex Detection of Low Abundance Point Mutations," *J. Mol. Biol.* 292:251–62 (1999) ("Gerry") (attached hereto as Exhibit 6) demonstrates the simultaneous detection of seven of the most common point mutations in the K-*ras* gene that are involved in colorectal cancer using My Array Design coupled to an LDR assay (Gerry at abstract and p. 255, para. bridging col. 1 and 2). LDR probe sets comprising an allele-specific probe with an addressable array portion (also referred to as "zip code") and a common probe having a fluorescent reporter label were designed to detect the seven mutations in nine individual DNA samples obtained from cell lines or paraffin-embedded tumor tissue (*id.* at p. 255, col. 2, para. 2 and Table 3). Following LDR, the ligated, fluorescently labeled LDR products were hybridized to an addressable DNA array containing capture oligonucleotides complementary to the addressable array sequences of the LDR products (*id.*).

41.   Using this method all K-*ras* mutations in the tumor and cell line DNA were correctly identified without the generation of false-positive or negative signals (*id.* at p. 256, para. bridging col. 1 and 2). To determine the limit of detection of low level mutations in wild-type DNA (*i.e.*, assay sensitivity), mutant DNA was diluted in wild-type DNA in ratios ranging from 1:20 to 1:500 (*id.* at p. 257, col. 2, para. 2). As shown in Figure 5 of Gerry, positive hybridization signal was quantifiable at a dilution of 1:200 with a signal-to-noise ratio of 2:1 (*id.*). These results confirmed the utility of My Array Design for detecting multiple nucleotide polymorphisms that are present in less than 1% of the total DNA (*id.*). We have subsequently fabricated a polymer flow-through biochip assembly that consists of a continuous-flow LDR microchip and a microarray chip that is capable of detecting one K-*ras* mutant sequence in the presence of 100 normal sequences (Hashimoto et al., "Ligase

Detection Reaction/Hybridization Assays Using Three-Dimensional Microfluidic Networks for the Detection of Low Abundant DNA Point Mutations," *Anal Chem* 77:3243–3255 at abstract (2005) (attached hereto as Exhibit 7)).

42. In addition to single nucleotide substitution mutations, many cancers involve small nucleotide insertions and deletions which result in frameshift mutations. For example, a number of small insertions and deletions are found within the BRCA1 and BRCA2 genes that are associated with inherited breast and ovarian cancer. A number of these insertion and deletion mutations are refractory to detection by direct hybridization array approaches, requiring the development of an alternative method. As described below, our technology is sensitive enough to detect sporadic mutations directly from tumor tissue within the p53 gene, which is involved in nearly half of all human cancers.

43. Favis et al., "Universal DNA Array Detection of Small Insertions and Deletion in BRCA1 and BRCA2," *Nat. Biotech.* 18:561–564 (2000) ("Favis") (attached hereto as Exhibit 8), demonstrates the capacity of My Array Design to reliably and reproducibly detect small nucleotide insertions and deletions using the BRCA1 and BRCA2 genes as a model system. As shown in Figure 1 of Favis, the method of detecting insertion and deletion mutations coupled a multiplex PCR step to LDR and My Array Design. This approach reproducibly detected both insertion and deletion mutations in BRCA1 and BRCA2 (*i.e.*, BRCA1 185delAG; BRCA1 5382insC; and BRCA2 6174delT). No cross-hybridization was detected, supporting the specificity of the method, and the reproducibility of the results were confirmed using a gel-based method (*id.* at p. 563, col. 2, para. 2). Further, even the presence of mutations in pooled samples was detected.

44. p53 mutations are observed in approximately one-half of all human cancers. My Array Design, when applied to the detection of p53 mutational status of clinical biopsy samples containing <5% tumor cells, was able to detect all mutations that were detected by direct sequencing and a yeast functional assay (Fouquet et al., "Rapid and Sensitive p53 Alteration Analysis in Biopsies from Lung Cancer Patients Using a Functional Assay and A Universal Oligonucleotide Array: A Prospective Study," *Clin Cancer Res* 10:3479–3489 at abstract and p. 3483, col. 2, para. 2 (2004) (attached hereto as Exhibit 9)). This approach was also used to detect 58 different p53 mutations in undissected colon tumor

DNA samples (Favis et al., "Harmonized Microarray/Mutation Scanning Analysis of TP53 Mutations in Undissected Colorectal Tumors," *Human Mutation* 24:63–75 (2004) (attached hereto as Exhibit 10)).

45.   An important feature of My Array Design is that it is not one-dimensional in it diagnostic utility. In addition to being a highly sensitive and robust method for detecting single base substitutions, insertions, and deletions involved in cancer development and progression, the method of coupling LDR to My Array Design has been successfully applied to the determination of promoter methylation status (Cheng et al., "Multiplexed Profiling of Candidate Genes for CpG Island Methylation Status Using a Flexible PCR/LDR/Universal Array Assay," *Genome Research* 16(2):282–9 at abstract (2006) (attached hereto as Exhibit 11). DNA methylation in CpG islands is associated with transcriptional silencing, and the ability to accurately determine cytosine methylation status in promoter CpG dinucleotides provides diagnostic and prognostic value for many human cancers. My Array Design demonstrated the ability to clearly distinguish different levels of methylation at 75 independent CpG dinucleotides in the promoter regions of 15 tumor suppressor genes (*id.*). When compared with an independent pyrosequencing method at a single promoter, the two approaches gave good correlation. In a study using 15 promoter regions and seven blinded tumor cell lines, our technology was capable of distinguishing methylation profiles that identified cancer cell lines derived from the same origins (*id.*). Further, our approach has the sensitivity required to detect the presence of methylation at 0.5% without selective PCR amplification, and at 0.05% with methyl-specific PCR amplification. This would correspond to identifying one tumor cell in 200 normal cells, or one tumor cell in 2,000 normal cells, respectively. This level of sensitivity holds the promise for early detection of colon cancer in DNA isolated from stool or serum.

Infectious Disease

46.   My Array Design can be utilized for identifying and distinguishing infectious agents (*e.g.*, bacterial, viral, and fungal) in many areas of biomedical science, including health care, biological defense, and environmental monitoring. The detection and identification of infectious agents must be highly sensitive and specific to distinguish closely related species or serotypes whose genomic sequences in specific regions differ at only a few nucleotide positions.

47.   Das et al., "Detection and Serotyping of Dengue Virus in Serum Samples by Multiplex Reverse Transcriptase PCR-Ligase Detection Reaction Assay," *J. Clin. Microbiol.* 46(10):3276–84 (2008) ("Das") (attached hereto as Exhibit 12) demonstrates the simultaneous serotyping and genotyping of dengue virus (DENV) in viral cultures and patient samples by coupling PCR based amplification to LDR and My Array Design. The assay accurately identified and serotyped DENV in 350 archived acute-phase serum samples, demonstrating 98.7% sensitivity and 98.4% specificity for detection (*id.* at p. 3280, col 2, para. 2). The detection limit for the assay ranged from 0.004 to 0.7 plaque forming units (PFU)/ reaction, comparable to those reported for other techniques (*id.* at para. bridging pp. 3282–83). The assay was highly specific for the detection of DENV with no cross reactivity to seven other similar flavivirus (*id.* at p 3281, col. 1, para. 2 and Figure 2). We have also employed this assay for the successful identification of West Nile viral strains, which also exhibit considerable genomic diversity, in clinical samples (*see* Rondini et al, "Development of Multiplex PCR-Ligase Detection Reaction Assay for Detection of West Nile Virus," *J. Clin. Microbiol.* 46:2269–79 (2008), attached hereto as Exhibit 13)

Conclusion:

48.   There are numerous advantages afforded by My Array Design. Direct hybridization arrays were designed to detect single nucleotide polymorphisms, with discrimination based on hybridization of target sequences to perfectly matched or mismatched probes. However, as discussed *supra*, the utility of direct hybridization methods to accurately and reproducibly detect and discriminate even these single nucleotide variations is questionable. The extent of sequence similarity between mismatch and perfect match probe sequences enables cross-hybridization and leads to the generation of both false-positive and false-negative signals. The findings of Naiser and Pozhitkov suggest that the problem of cross-hybridization is only further compounded by the unpredictable nature of oligonucleotide probe-target hybridization. Direct hybridization methods are not well suited for the detection of most other types of nucleic acid sequence variations. In fact, in most cases, the ability to detect insertion/deletion mutations has proven intractable.

49.   In contrast to direct hybridization methods, My Array Design has demonstrated the ability to detect insertion/deletion mutations, mononucleotide and

dinucleotide repeats, and even methylation of CpG islands. In addition, it can be used to identify and quantify splice site changes, quantify RNA levels for gene expression profiling, and determine DNA copy levels changes, loss of heterozygosity, and SNPs for genome-wide association studies.

50.   A significant advantage of the My Array Design is that it relies on divergent capture-specific probe sequences, designed to differ in sequence by at least 25%, yet have similar melting temperatures. The result: cross-hybridization is minimized, if not eliminated, even under uniform hybridization conditions. The composite probes and products, which contain a target specific portion and a capture specific portion allow for accurate target identification and discrimination of closely spaced and overlapping mutations, including small insertions and deletions without generating false-positive or false-negative signals. In addition, the method has proven to be highly sensitive, capable of detecting low abundance mutations in heterogenous clinical samples. This sensitivity permits early disease detection, which can be critical for a good disease prognosis. The ability to use the method to detect promoter methylation silencing of tumor suppressor genes helps predict disease outcome and guide cancer treatment.

51.   I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Date:_March 27, 2010__

Francis Barany

Francis Barany, Ph.D.

.

Exhibit 1:   Francis Barany Curriculum Vitae

CURRICULUM VITAE

FRANCIS BARANY

Business Address:
Department of Microbiology & Immunology
   Box 62, Room B-406
Weill Cornell Medical College
1300 York Avenue
New York, NY 10065
Tel   212.746.6509 or 6507 or 6524
Fax   212.746.8104
e.mail barany@med.cornell.edu

Home Address:
450 East 63rd St. Apt. #12C
New York, NY 10065
Tel   212.371.6158
Mobile 917.957.3976

Personal Data:

Date of Birth:	April 4, 1957
Place of Birth:	Afula, Israel; Emigrated to U.S. in 1960
Citizenship:	United States, 1965
Marital Status:	Married (Rachel Conescu, Nov. 1991)
Children:	Isabelle (Born November 19, 1993)
	Lillian (Born January 30, 1996)

Education:

University of Illinois at Chicago Circle, 1974-1976
B.A. in Chemistry, 1976

The Rockefeller University, New York, 1976-1981
Ph.D. in Microbiology, 1981.   Thesis: Plasmid transformation in *Streptococcus pneumoniae*. Research Advisor: Alexander Tomasz

Employment:

1981-1982   The Rockefeller University, New York
              Postdoctoral Fellow; Alexander Tomasz

1982-1985   The Johns Hopkins University School of Medicine, Baltimore, MD
              Postdoctoral Fellow; Hamilton O. Smith

1985-1990   Cornell University Medical College, New York
              Assistant Professor, Dept. of Microbiology

1990-1995   Cornell University Medical College, New York
              Associate Professor, Dept. of Microbiology, Tenure Granted, June 1993

1986-2003   The Rockefeller University, New York
              Adjunct Assistant, Associate, Full Professor

1994-2007   Strang Cancer Prevention Center, New York
              Director of Mutation Research

1995-present Cornell University Medical College, New York
              Professor, Dept. of Microbiology
              Professor, Program of Biochemistry and Structural Biology

Academic Honors:

Westinghouse Science Talent Search, National Finalist, 1974
B.B. Freund Award for excellence in Chemistry, 1975
The American Institute of Chemists Award, 1976
National Science Foundation Predoctoral Fellowship, 1976-1979
Rockefeller University Graduate Fellowship, 1980-1981
Andrew W. Mellon Postdoctoral Fellow, 1981-1982
Helen Hay Whitney Fellow, 1982-1985
Cornell Scholar in Biomedical Sciences, 1985-1988
Hirschl/Monique Weill-Caulier Career Scientist Award, 1992-1997
Mayent/Rothschild Visiting Professor, Institut Curie, Paris, August 2000
Technology Innovator Award, Bioarrays 2003-New York Meeting.
Medical Diagnostics Research leader, Scientific American 50, 2004
Weill Cornell Medical College Award for Teaching Excellence, 2007

Recent Grants:

Development of a New Class of Drugs: Coferons
    Coferon Inc.
    2010-2011 $1,840,000 direct cost for Barany Lab
Promoter Methylation of LRAT Gene
    Exact Sciences
    2009-2011 $420,000 direct cost for Barany Lab
Identification Of New Therapeutic Targets And Markers Of Colon Cancer
    Ludwig Institute for Cancer Research
    2007-2009 $850,000 direct cost for Barany Lab
Multiplexed Detection of Food and Waterborne Pathogens – NIAID 1U01 AI075470-01
    2008-2013. F. Barany, Project Director
    $2,736,876 direct cost for Barany Lab/Cornell
Identification of Compounds that Kill Non-replicating Mtb and their targets
    Bill and Melinda Gates Foundation
    2007-2008. C. Nathan, Project Director
    $ 1,390,893 total direct cost, $139,810 direct cost for Barany Lab
Multiplexed Detection of Bioterror Agents – NIAID 1 UC1 A1062579-01
    2004-2007. F. Barany, Project Director
    $2,720,023 direct cost for Barany Lab/Cornell
Array Technology, Mutation Scanning Technology, and its use in Diagnosis.
    Applied Biosystems (yearly renewal)
    2004-2005 $592,593 direct cost + $400,000 equipment for Barany Lab
New Methods for Cancer Detection -- NCI P01-CA65930-05
    2002-2006. F. Barany, Program Director
    $10,357,226 total direct cost, $3,636,053 direct cost for Barany Lab
Array Technology, Mutation Scanning Technology, and its use in Diagnosis.
    Applied Biosystems

2001-2004 $1,820,119 direct cost for Barany Lab
Identifying Genome Changes in Cancer Development -- NCI
Priority score: 136 = 2.5 percentile
1998-2001. $1,037,911 direct cost for Barany Lab
DNA arrays for cancer detection. – Applied Biosystems/Perkin Elmer
1998-2001. $975,000 direct cost for Barany Lab
New Methods for Cancer Detection -- NCI P01-CA65930-01A1
1997-2002. $2,647,728 direct cost for Barany Lab
Germline Alterations of Tumor Suppressor Genes in New York Cancer Patients
1999-2001. $125,000 direct cost for Barany Lab
Mechanism of Restriction Endonuclease Action -- NIH RO1 GM41337
1994-1998. $582,761 direct cost.
Micro-automated thermal cycler --NIST
Joint Venture with 2 other universities, and 2 industrial labs.
1995-1998. $270,880 direct cost for Barany Lab.
Ligation Amplification Technology -- Applied Biosystems Inc.
1992-1997. $663,100 direct cost.

Research Interests:

DNA uptake and integration mechanisms - transport of biological macromolecules
across membrane barriers
Design and synthesis of oligonucleotide linkers - in vitro mutagenesis to obtain
proteins with specifically altered structures
Understanding the mechanism of sequence-specific DNA recognition by proteins
Protein engineering
Identification and understanding the molecular basis of genetic diseases and cancers
Molecular profiling of mutational, epigenetic and LOH changes associated with cancers
Single molecule detection for profiling DNA and protein changes associated with tumor
development
High-throughput, small molecule screening of compounds that kill non-replicating cells
Molecular identification of early cancer
Molecular signatures that predict outcome and guide treatment of cancers
Coferons; Design and synthesis of orally active drug molecules that can cross
membrane barriers, and once inside can combine to interfere with target activity.

Research Experience and Training:

Characterization of Venus' flytrap proteases (1971-1974)
Organic chemistry (J. Kagan, 1975-1976)
Organic photochemistry, NMR, IR, GC, and Mass Spec (S. Wolff & W.C. Agosta, 1977)
Photosystem II biochemistry of thermophilic blue green algae (D.C. Mauzerall, 1978)
Pneumococcal transformation (A. Tomasz, 1978-1982)
Plasmid & Phage molecular biology (P. Model & N.D. Zinder, 1980-1982)
Haemophilus transformation (H.O. Smith, 1982-1983)
Oligonucleotide linkers for in vitro mutagenesis (1983-1992)
Characterization of isoschizomeric *Taq*I and *Tth*HB8 I restriction endonucleases (1985-1998)
Cloning and use of thermostable ligase for genetic disease and cancer detection (1989-present)
Design and use of addressable DNA arrays (1993-present)

Cloning and use of thermostable Endonuclease V for scanning unknown mutations in cancers (1998-present)
Representational genomic amplification and sequencing for large-scale DNA array analysis of SNPs and allelic imballance in cancers (1999-present)
Methylation profiling of tumors (2000-present)
Single Molecule detection technology (2002-present)
Multiplexed detection of pathogens (2003-present)

Professional Activities:

Chairman of Partnerships for Biodefense Food- and Water-borne Diseases Study Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health January, 2010
Chairman of Partnerships for Point of Care (POC) Diagnostic Technologies for Nontraditional Health Care Settings Study Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health October, 2008
Co-founder of the New York State Cancer Initiative working committee, 2007
Member of the Ensemble Scientific Advisory Board 2007
Member of Scientific Advisory Board, Center for BioModular Multi-scale Systems, LA, Oct. 2005
Chairman of Innovative Technologies for the Molecular Analysis of Cancer Study Section, National Cancer Institute, National Institutes of Health November, 2003, July, 2004
Member of study section reviewing PO1, National Institutes of Health, March 2002, November 2002
Member of Site Visit Team reviewing Jackson Labs, National Cancer Institute, National Institutes of Health, February 2001
Member of SBIR study section, National Institutes of Health, July 2000
Co-organizer, 1998 FASEB Conference "Nucleic Acid Enzymes: Mechanisms and Diseases."
Member of Innovative Technologies for the Molecular Analysis of Cancer Study Section, National Cancer Institute, National Institutes of Health November, 1998, July 1999, July 2000, March 2002
Member of Novel Technologies for Evaluation of Molecular Alterations in Tissue, and Technologies for Generation of Full-Length cDNA Libraries, Special Study
Sections, National Cancer Institute, National Institutes of Health, July 1997, March 1998.
Ad-hoc member of Developmental Diagnostics Working Group, National Cancer Institute, National Institutes of Health, July 1997.
Member of Advanced Diagnostics for Pathogens Study Section, Defense
Advanced ResearchProjects Agency, June and September 1997. August, 1998.
Ad-hoc member of Human Genome Study Section, National Institutes of Health November 1994 and June 1995.
Site Visit of National Cancer Institute Early Detection Research Network, June 1995.
Member of Academic Medicine Development Corporation New York Cancer Project Consortia Group (1997-Present)
Editor of Gene (1987- 1995)
Editorial Advisory Board Member of Gene (1996)
Scientific Advisory Board of Amplicon (1995-1998)
Referee of research articles submitted for Science, Proceedings of the National Academy of Sciences, EMBO Journal, Gene, J. Bacteriology, Biochemistry, and Nucleic Acids Research.
Reviewer for NSF grants

Professional Societies:

        Phi Eta Sigma, 1974
        Sigma Xi, 1980
        American Society for Microbiology, 1990
        American Society for Biochemistry and Molecular Biology, 1990

Patents:

The Barany Laboratory patents (27) have been responsible for more than 10% of the issued patents from Weill Cornell Medical College. The Barany Laboratory patents and intellectual property have generated over $28 million in NIH Grants, NIST Grant, Industrial Sponsored Research Grants, over $6 million in royalties to Weill-Cornell, and over $1.3 billion in sales or value to biotechnology companies.

1. Barany, F. Six base oligonucleotide linkers and methods for their use. Licensed to Pharmacia P-L Biochemicals, WI. (U.S. Patent #4,719,179; issued May 1988).

2. Barany, F., Zebala, J. Nickerson, D., Kaiser, R., & Hood L. Thermostable ligase mediatedDNA amplification system for the detection of genetic diseases. Licensed to Applied Biosystems/Perkin Elmer Inc., Foster City, CA. Thermostable ligase licensed to Roche Molecular Systems, Alameda, CA and New England Biolabs, Beverly, MA (U.S. Patent # 5,494,810, issued February, 1996; U.S. Patent # 5,830,711, issued November, 1998; and, U.S. patent #6,054,564, issued April, 2000).

3. Barany, F., Barany, G., Hammer, R.P., Kempe, M., Blok, H., & Zirvi, M. Detection of nucleic acid sequence differences using the ligase detection reaction with addressable arrays. Licensed to Applied Biosystems/Perkin Elmer Inc., Foster City, CA. (U.S. patent # 6,852,487; issued February, 2005; and U.S. patent # 7,083,917; issued August, 2006).

4. Barany F., Belgrader, P., & Lubin, M. Detection of nucleic acid sequence differences using coupled ligase detection and polymerase chain reactions. Licensed to Applied Biosystems/Perkin Elmer Inc., Foster City, CA (U.S. patent #6,027,889, issued February, 2000; U.S. patent #6,268,148, issued July, 2001; U.S. patent # 6,797,470, issued September 2004;

5. Barany F., Lubin, M., Barany, G., & Hammer, R.P. Detection of nucleic acid sequence differences using coupled ligase detection and polymerase chain reactions. Licensed to Applied Biosystems/Perkin Elmer Inc., Foster City, CA U.S. patent # 7,097,980, issued August 2006; U.S. patent # 7,166,434, issued January 2007; U.S. patent # 7,312,039, issued December 2007, U.S. patent # 7,320,865, issued January 2008, U.S. patent # 7,332,285, issued February 2008, U.S. patent # 7,364,858, issued April 2008, U.S. patent # 7,429,453, issued September 2008, U.S. patent # 7,556,924, issued July 2009).

6. Barany, F., Luo, J., Khanna, M., & Bergstrom, D. High fidelity detection of nucleic acid differences by ligase detection reaction. Licensed to Applied Biosystems/Perkin Elmer Inc., Foster City, CA. (U.S. patent # 6,312,892; issued November, 2001; U.S. patent # 7,244,831; issued July, 2007).

7. Barany, F., Luo, J., & Bergstrom, D. Thermostable DNA ligase mutants. Licensed to Applied Biosystems/Perkin Elmer Inc., Foster City, CA. (U.S. patent #6,576,453; issued June, 2003).

8. Barany, F., Cao, W., Tong, J. High Fidelity thermostable ligase and uses thereof. Licensed to Applied Biosystems, Foster City, CA (U.S. patent # 6,949,370, issued September 2005).

9. Barany F., Zirvi, M., Gerry, N., Paty, P. Accelerating identification of single nucleotide polymorphisms and alignment of clones in genomic sequencing. Licensed to Applied Biosystems, Foster City, CA (pending, U.S. patent office, submitted, January, 1999).

10. Barany, F., Gerry, N., Witowski, N, Day, J., Hammer, R.P., Barany, G. Detection of nucleic acid sequence differences using the ligase detection reaction with addressable arrays. Licensed to Applied Biosystems, Foster City, CA (U.S. patent #6,506,594, January 2003).

11. Barany, F., Day, J., Hammer, R.P., Bergstrom, D. Coupled polymerase chain reaction-restriction endonclease digestion- ligase detection reaction process. Licensed to Applied Biosystems, Foster City, CA (U.S. patent #7,014,994, issued March 2006).

12. Barany F., Liu, Z, Kirk, B., Zirvi, M., Gerry, N., Paty, P. Accelerating identification of single nucleotide polymorphisms and alignment of clones in genomic sequencing. (U.S. patent #6,534,293, March 2003).

13. Barany F., Zirvi, M., Gerry, N., Favis, R., Kliman, R. Method of designing an addressable array suitable for detection of nucleic acid sequence differences using the ligase detection reaction with addressable arrays (U.S. patent # 7,455,965, issued November 2008).

14. Barany, F., Cao, W., J. Huang, J. Lu. Detection of nucleic acid differences using combined endonuclease cleavage and ligase reactions. Licensed to Applied Biosystems, Foster City, CA (U.S. patent # 7,198,894, issued April 2007).

15. Barany, F., Turner, D., Pingle, M., and Pincas, H. Methods for identifying target nucleic acid molecules. (pending, U.S. patent office, submitted, September, 2003).

16. Barany, F., Cheng, Y-W., Shawber, C. Method for detection of promoter methylation status. (U.S. patent # 7,358,048, issued April 2008).

17. Barany, F., Huang, J., Pincas, H., and Lao, K. Detection of nucleic acid differences using endonuclease cleavage reactions and capillary electrophoresis or microarrays (submitted, July, 2004).

18. Barany, F., Cheng, Y.-W., Paty, P. and Notterman, D. Use of Lecithin-Retinol Acyl Transferase gene promoter methylation in evaluating the cancer state of a subject (submitted, December, 2006).

19. Barany, F., Pingle, M., and Bergstrom D. Detection of target nucleic acid sequences using fluorescence resonance energy transfer (submitted, April, 2008).

20. Barany, F., Pingle, M., and Bergstrom D. Coferons and methods of making and using them (submitted, April, 2008).

21. Barany, F., Zirvi, M., Giardina, S., Bacolod, M., Cheng, Y.-W., Parker, O., Schemmann, G., Notterman, D., and Paty, P. Methods for predicting disease outcome for patients with colon cancer. (submitted, October, 2008).

22. Barany, F., Pingle, M., Giardina, S. and Bergstrom D. Coferons and methods of making and using them (submitted, April, 2009).

23. Barany, F., Pingle, M., Bergstrom D and Arnold, L. Coferons and methods of making and using them (submitted, October, 2009).

BIBLIOGRAPHY (ARTICLES)

1. Barany, F., Wolff, S. & Agosta, W.C. (1978) Novel temperature- dependent photochemical rearrangement of Citral. J. Amer. Chem. Soc. 100:1946-1948.

2. Robertson, H.D. & Barany, F. (1979) Enzymes and mechanisms in RNA processing. In Gene Function 51:285-295. Febs Meeting, 1978, Eds. S. Rosenthal et al. Pergamon Press, New York.

3. Wolff, S., Barany, F. & Agosta, W.C. (1980) Novel photochemical rearrangements of Citral and related compounds at elevated temperatures. J. Amer. Chem. Soc. 102:2378-2386.

4. Barany, F. & Tomasz, A. (1980) Genetic transformation of *Streptococcus pneumoniae* by heterologous plasmid deoxyribonucleic acid. J. Bacteriol. 144:698-709.

5. Barany, F. & Boeke, J.D. (1982) Plasmid exchange between *Streptococcus pneumoniae* and *Escherichia coli*. In Genetic Exchange: A celebration and a new generation. Eds. Streips, U.A., Goodgal, S., Guild, W. & Wilson, G.A. Marcel Dekker, Inc. New York, pp 27-42.

6. Barany, F., Boeke, J.D. & Tomasz, A. (1982) Staphylococcal plasmids which replicate and express erythromycin resistance in both *Streptococcus pneumoniae* and *Escherichia coli*. Proc. Natl. Acad. Sci. USA 79:2991-2995.

7. Barany, F. (1982) Transformation of *Streptococcus pneumoniae* by single-stranded plasmid-phage hybrid DNA. In Microbiology. Ed. Schlessinger, D. pp. 125-129.

8. Barany, F. & Boeke, J.D. (1983) Genetic transformation of *Streptococcus pneumoniae* by DNA cloned into the single- stranded bacteriophage fl. J. Bacteriol. 153:200-210.

9. Kahn, M.E., Barany, F. & Smith, H.O (1983) Transformasomes: Specialized membranous structures that protect DNA during *Haemophilus* transformation. Proc. Natl. Acad. Sci. USA 80:6927-6931.

10. Barany, F., Kahn, M.E. & Smith, H.O. (1983) Directional transport and integration of donor DNA in *Haemophilus influenzae* transformation. Proc. Natl. Acad. Sci. USA 80: 7274-7278.

11. Barany, F. & Kahn, M.E. (1985) A comparison between transformation mechanisms of *Haemophilus parainfluenzae* and *Haemophilus influenzae*. J. Bacteriol. 161:72-79.

12. Barany, F. (1985) Two codon insertion mutagenesis of plasmid genes using single-stranded hexameric oligonucleotides. Proc. Natl. Acad. Sci. USA 82:4202-4206.

13. Barany, F. (1985) Single-stranded hexameric linkers: A system for in-phase insertion mutagenesis and protein engineering. Gene 37:111-123.

14. Barany, F. (1987) A genetic system for isolation and characterization of *Taq*I restriction endonuclease mutants. Gene 56: 13-27.

15. Barany, F. (1988) The *Taq*I star reaction: strand preferences reveal hydrogen bond donor and acceptor sites in canonical sequence recognition. Gene 65: 149-165.

16. Barany, F. (1988) Overproduction, purification, and crystallization of *Taq*I restriction endonuclease. Gene 65: 167-177.

17. Barany, F. (1988) Procedures for linker insertion mutagenesis and use of new kanamycin resistance cassettes. DNA and Protein Engineering Techniques 1: 29-35.

18. Barany, F. (1988) How *Taq*I endonuclease recognizes its cognate sequence. Gene 74:63-65.

19. Glushka, J., Barany, F., & Cowburn, D. (1989) Observation of arginyl-deoxyoligonucleotide interactions in *Taq*I endonuclease by detection of specific 1H NMR signals from 140kD [Nn1, Nn2, ^{15}N Arg]*Taq*I/oligomer complexes. Biochem. & Biophys. Res. Commun. 164:88-93.

20. Barany, F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase. Proc. Natl. Acad. Sci. USA, 88:189-193.

21. Zebala, J., Mayer, A., & Barany, F. (1991) Correlation between the homologies of *Tth*HB8 and *Taq*I isoschizomers and insertion mutants in the *Taq*I endonuclease. In Proceedings of the International Symposium on Site Directed Mutagenesis and Protein Engineering. M.R. El-Gewely, ed. Elsevier Science Publishers B.V., Amsterdam, pp. 133-140.

22. Zebala, J., & Barany, F. (1991) Mapping catalytically important regions in ß-lactamase using two codon insertion mutagenesis. In Proceedings of the International Symposium on Site Directed Mutagenesis and Protein Engineering. M.R. El-Gewely, ed. Elsevier Science Publishers B.V., Amsterdam, pp. 189-192

23. Zebala, J., & Barany, F. (1991) Mapping catalytically important regions of an enzyme using two-codon insertion mutagenesis: a case study correlating ß-lactamase mutants with the three dimensional structure. Gene 100:51-57.

24. Barany, F. (1991) The ligase chain reaction (LCR) in a PCR world. (Invited Review) PCR Methods and Applications 1: 5-16.

25. Barany, F. and Gelfand, D. (1991) Cloning, overexpression, and nucleotide sequence of a thermostable DNA ligase-encoding gene. Gene. 109:1-11.

26. Barany, F., Danzitz, M., Zebala, J. and Mayer, A. (1992) Cloning and sequencing of genes encoding the *Tth*HB8I restriction and modification enzymes; comparison with the isoschizomeric *Taq*I enzymes. Gene. 112:3-12.

27. Barany, F. and Zebala, J. (1992) Correlation between insertion mutant activities and amino acid identities of the *Taq*I and *Tth*HB8I restriction endonucleases. Gene. 112:13-20.

28. Barany, F., Slatko, B., Danzitz, M., Cowburn, D., Schildkraut, I. and Wilson, G. (1992) The corrected nucleotide sequence of the *Taq*I restriction and modification enzymes reveals a thirteen-codon overlap. Gene. 112:91-95.

29. Zebala, J., Choi, J. and Barany, F. (1992) Characterization of steady state, single-turnover and binding kinetics of the *Taq*I restriction endonuclease. J. Biol. Chem. 267: 8097-8105.

30. Zebala, J., Choi, J., Trainor, G. and Barany, F. (1992) DNA recognition of base-analogue and chemically modified substrates by the *Taq*I restriction endonuclease. J. Biol. Chem. 267:8106-8116.

31. Wiedman, M., Czajka, J., Barany, F. and Batt, C. (1992) Discrimination of *Listeria monocytogenes* from other *Listeria* species by ligase chain reaction. Appl. Environ. Microbiol. 58:3443-3447.

32. Prchal, J.T., Guan, Y.L, Prchal, J.F. and Barany, F. (1993) Transcriptional analysis of the active X-chromosome in normal and clonal hematopoiesis. Blood 81: 269-271.

33. Feero, W.T., Wang, J., Barany, F., Zhou, J., Todorovic, S.M., Conwit, R., Galloway, G., Hausmanowa-Petrusewicz, I., Fidzianska, A., Arahata, K., Wessel, H.B., Wadelius, C., Marks, H.G., Hartlage, P., Hayakawa, H., and Hoffman, E.P. (1993) Hyperkalemic Periodic Paralysis: Rapid molecular diagnosis and relationship of genotype to phenotype in 12 families. Neurology 43:668-673.

34. Wang, J., Zhou, J., Todorovic, S.M., Feero, W.G., Barany, F., Conwit, R., Hausmanowa-Petrusewicz, I., Fidzianska, A., Arahata, K., Wessel, H.B., Sillen, A., Hayakawa, H., and Hoffman, E.P. (1993) Molecular genetic and genetic correlations in sodium channelopathies: Lack of founder effect and evidence of a second gene. Am. J. Hum. Genet. 52: 1074-1084.

35. Zebala, J., and Barany, F. (1993) Implications for the ligase chain reaction in gastroenterology. J. Clin. Gastroenterol. 17:171-175.

36. Wiedman, M., Barany, F. and Batt, C. (1993) Detection of *Listeria monocytogenes* with a non-isotopic polymerase chain reaction-coupled ligase chain reaction assay. Appl. Environ. Microbiol. 59:2743-2745.

37. Wiedman, M., Wilson,W., Czajka, J., Luo, J., Barany, F., and Batt, C. (1994) Ligase chain reaction (LCR)- Overview and applications. PCR Methods Appl. 3:S51-S64.

38. Ruiz-Opaz, N., Barany, F., Hirayama, K., and Herrera, V. (1994) Confirmation of mutant α1 Na,K-ATPase gene and transcript in Dahl salt-sensitive/JR rats. Hypertension 24:260-270.

39. Wiedman, M., Wilson, W., Czajka, J., Barany, F., and Batt, C (1994) Ligase mediated detection techniques. In *Methods in DNA Amplification*, Rolfs, A. et al. ed. Plenum Press, NY, p83-92.

40. Mayer, A., and Barany, F. (1994) Interaction of *Taq*I endonuclease with the phosphate backbone: Effects of stereospecific phosphate modification. J. Biol. Chem. 269:29067-29076.

41. Mayer, A., and Barany, F. (1995) Photoaffinity crosslinking of *Taq*I endonuclease using an aryl azide linked to the phosphate backbone. Gene 153:1-8.

42. Cao, W. , Mayer, A., and Barany, F. (1995) Stringent and relaxed specificity of *Taq*I endonuclease: interactions with metal cofactors and DNA sequences. Biochemistry 34:2276-2283.

43. Zebala, J. and Barany, F. (1995) Detection of Leber's Hereditary Optic Neuropathy by nonradioactive-LCR. In *PCR Strategies*, M.A. Innis, D.H. Gelfand, and J.J. Sninsky, eds., Academic Press, San Diego, pp 335-346.

44. Wiedman, M., Barany, F. and Batt, C. (1995) Detection of *Listeria monocytogenes* by PCR-coupled ligase chain reaction (LCR). In *PCR Strategies*, M.A. Innis, D.H. Gelfand, and J.J. Sninsky, eds., Academic Press, San Diego, pp 347-361.

45. Wiedman, M., Barany, F. and Batt, C. (1995) Ligase Chain Reaction. In *PCR Primer, A Laboratory Manual* M.A. Innis, C.W. Dieffenbach and G.S. Dveksler, eds., Cold Spring Harbor Laboratory Press, New York, pp 631-652.

46. Day, D., Speiser, P.W, White, P.C., and Barany, F. (1995) Detection of steroid 21 hydroxylase alleles using gene -specific PCR and a multiplexed ligation detection reaction. Genomics. 29:152-162.

47. Belgrader, P.. Marino, M., Lubin, M., and Barany, F. (1996) A Multiplex PCR-Ligase Detection Reaction Assay for Human Identity Testing. Genome Science and Technology. 1:77-87.

48. Luo, J. and Barany, F. (1996) Identification of essential residues in *Thermus thermophilus* DNA ligase. Nucleic Acids Research. 24:3079-3085.

49. Luo, J., Bergstrom, D., and Barany, F. (1996) Improving the fidelity of *Thermus thermophilus* DNA ligase. Nucleic Acids Research. 24:3071-3078.

50. Day, D.J., Speiser, P.W., Schulze, E., Bettendorf, M., Fitness, J., Barany, F. and White, P.C. (1996) Identification of non-amplifying CYP21 genes when using PCR based diagnosis of 21-hydroxylase deficiency in congenital adrenal hyperplasia (CAH) affected pedigrees. Human Molecular Genetics 5:2039-2048.

51. Cao, W., Lu, J., and Barany, F. (1997) Nucleotide sequences and gene organization of *Taq*I endonuclease isoschizomers from *Thermus sp.* SM32 and *Thermus filiformis* Tok6A1 Gene. 197:205-214.

52. Cao, W., Lu, J., Welch, S. G., Williams, R. A. D. and Barany. F. (1998) Cloning and thermostability of *Taq*I endonuclease isoschizomers from *Thermus sp.* SM32 and *Thermus filiformis.* Tok6A1. Biochemistry J. 333:425-431.

53. Khanna, M., Park, P., Zirvi, M., Cao, W., Picon, A., Day, J., Paty, P. and Barany, F. 1999. Multiplex PCR/LDR for detection of K-*ras* mutations in primary colon tumors. Oncogene 18:27-38.

54. Khanna, M., Cao, W., Zirvi, M., Paty, P. and Barany, F. 1999. Ligase Detection Reaction for Identification of Low Abundance Mutations. Clinical Biochemistry 32:287-290.

55. Cao, W. and Barany. F. 1998. Identification of *Taq*I endonuclease active site by Fe^{2+} mediated oxidative cleavage. J. Biol. Chem. 273: 33002-33010.

56. Tong, J., Cao, W. and Barany, F. 1999. Biochemical properties of a high fidelity DNA ligase from *Thermus* species AK16D. Nucleic Acids Research 27:788-794.

57. Day, J., Bergstrom, D. , Hammer, R. and Barany F. 1999. Nucleotide analogs facilitate base conversion with 3' mismatch primers. Nucleic Acids Research 27:1810-1818.

58. Day, J., Hammer, R., Bergstrom, D. and Barany F. 1999. Nucleotide analogs and new buffers improve a generalized method to enrich for low abundance mutations. Nucleic Acids Research 27:1819-1827.

59. Blumberg, D., Picon, A., Klimstra, D., Cohen, A., Barany, F., and Paty, P. Evaluation of K-ras Mutation Status as a Predictor of Lymph Node Metastases in T1/T2 Rectal Cancers. Diseases of Colon and Rectum, In Press, 3/15/99.

60. Gerry, N., Witowski, N., Day, J., Hammer, R., Barany, G. and Barany, F. 1999. Universal DNA array with polymerase chain reaction/ligase detection reaction (PCR/LDR) for multiplex detection of low abundance mutations. J. Mol. Biol. 292:251-262.

61. Zirvi, M., Nakayama, T., Newman, G., McCaffrey, T., Paty, P. and Barany, F. 1999. Ligase-based Detection of Mononucleotide Repeat Sequences. Nucleic Acids Research, 27: e40.

62. Zirvi, M., Bergstrom, D., Saurage, A. Hammer, R. and Barany F. 1999. Improved Fidelity of Thermostable Ligases for Microsatellite Repeats using Nucleoside Analogues. Nucleic Acids Research, 27:e41.

63. Favis, R. and Barany, F. 2000. Mutation Detection in K-ras, BrCA1, BrCA2 and p53 using Multiplex PCR/LDR and a Universal DNA Microarray. "Circulating Nucleic Acids in Plasma or Serum", eds. Philippe Anker and Maurice Stroun in Vol. 96 *Annals of the New York Academy of Sciences.* 906:39-43.

64. Tong, J., Barany, F, and Cao, W. 2000. Ligation Reaction Specificities of an NAD$^+$-dependent DNA Ligase from Hyperthermophile *Aquifex aeolicus.* Nucleic Acids Research, 28:1447-1454.

65. Favis, R., Day, J.P., Gerry, N., Narod, S.A., Paty, P. and Barany, F. 2000. Multiplex PCR/PCR/LDR detection of BRCA1 and BRCA2 small insertions and deletions using a Universal DNA array. Nature Biotechnology, 18:561-564.

66. Dong, S.M., Traverso, G., Johnson, C., Geng, L., Favis, R., Boynton, K., Hibi, K., Goodman, S., D'Allesio, M., Paty, P., Hamilton, S., Sidransky, D., Barany, F., Levin, B., Shuber, A., Kinzler, K., Vogelstein, B., and Jen, J. 2001 Detecting colorectal cancer in stool using multiple genetic targets. J. Natl. Cancer Inst. 93:858-865.

67. Huang, J., Lu, J., Barany, F., and Cao, W. 2001. Multiple Cleavage Activities of Endonuclease V from *Thermotoga maritima*: Recognition and Strand Nicking Mechanism. Biochemistry 40:8738-8748.

68. Huang, J., Lu, J., Barany, F., and Cao, W. 2002 Mutational Analysis of Endonuclease V from *Thermotoga maritima.* Biochemistry 41:8342-8350.

69. Huang, J., Kirk, B., Favis, R., Soussi, T., Paty, P., Cao, W. and Barany, F. 2002. An Endonuclease /Ligase based mutation scanning method especially suited for analysis of neoplastic tissue. Oncogene 21: 1909-1921.

70. Kirk, B.W., Feinsod, M, Favis, R., Kliman, R.M., and Barany, F. 2002. Single Nucleotide Polymorphism Seeking Long Term Association with Complex Disease. Invited review and cover article for Nucleic Acid Research. 30:3295-3311.

71. Wang Y., Vaidya B., Farquar H.D., Stryjewski W., Hammer R.P., McCarley R.L., Soper S.A., Cheng Y.W., and Barany F. 2003 Microarrays assembled in microfluidic chips fabricated from poly(methyl methacrylate) for the detection of low-abundant DNA mutations. Analytical Chemistry 75: 1130-1140.

72. Nathanson, D.R., Culliford, A.T. 4th, Shia, J., Chen, B., D'Alessio, M., Zeng, Z.S., Nash, G.M., Gerald, W., Barany, F., and Paty, P.B..2003 HER 2/neu expression and gene amplification in colon cancer. Int J Cancer. 105:796-802.

73. Wabuyele, M.B., Farquar, H., Stryjewski, W., Hammer, R.P., Soper, S.A., Cheng, Y.W., and Barany, F. 2003. Approaching Real-Time Molecular Diagnostics: Single-Pair Fluorescence Resonance Energy Transfer (spFRET) Detection for the Analysis of Low Abundant Point Mutations in K-ras Oncogenes. J Am Chem Soc. 125: 6937-6945.

74. Nash, G.M., Gimbe,l M., Shia, J., Culliford, A.T., Nathanson, D.R., Ndubuisi, M., Yamaguchi, Y., Zeng, Z.S., Barany, F., and Paty, P.B. 2003 Automated, multiplex assay for high-frequency microsatellite instability in colorectal cancer. J Clin Oncol. 21:3105-12.

75. Overholtzer, M., Rao, P.H., Favis, R., Lu, X.Y., Elowitz, M.B., Barany, F., Ladanyi, M., Gorlick, R., Levine, A.J. 2003 The presence of p53 mutations in human osteosarcomas correlates with high levels of genomic instability. Proc Natl Acad Sci U S A. 100:11547-52.

76. Fouquet, C., Antoine, M., Tisserand, P., Favis, R., Wislez, M., Como, F., Rabbe, N., Carette, M.F., Milleron, B., Barany, F., Cadranel, J., Zalcman, G. and Soussi, T. 2004. p53 sequence analysis in biopsies from a lung cancer patient using a versatile oligonucleotide array: A prospective study. Clinical Cancer Res. 10:3479-3489.

77. Thomas G., Sinville R., Sutton S., Farquar H., Hammer R.P., Soper S.A., Cheng Y.W., and Barany F. 2004. Capillary and microelectrophoretic separations of ligase detection reaction products produced from low-abundant point mutations in genomic DNA. Electrophoresis. 25:1668-1677.

78. Favis, R., Huang, J., Gerry, N., Culliford, A., Paty, P., Soussi, T. and Barany, F. 2004. Universal DNA Microarray Analysisof p53 Mutations in Undissected Colorectal Tumors. Human Mutation. 24:63-75.

79. Gimbel M, Nash G, Ndubuisi M, Wong WD, Barany F, and Paty PB. 2004. Braf mutations are associated with increased mortality in colorectal cancer. J Amer Coll Surgeons. 199(3):S92.

80. Lu, J., Tong, J., Feng, H, Huang, J., Afonso, C.L., Rock, D.L., Barany, F., and Cao, W. 2004. Unique ligation properties of eukaryotic NAD$^+$-dependent DNA ligase from *Melanoplus sanguinipes* entomopoxvirus. BBA 1701:37-48.

81. Pincas, H., Pingle, M., Huang, J., Lao, K., Paty, P.B., Friedman, A.M. and Barany, F. 2004. High sensitivity EndoV mutation scanning through real-time ligase proofreading. Nucleic Acids Research 32:e148.

82. Phelan CM, Dapic V, Tice B, Favis R, Kwan E, Barany F, Manoukian S, Radice P, van der Luijt RB, van Nesselrooij BP, Chenevix-Trench G, kConFab, Caldes T, de la Hoya M, Lindquist S, Tavtigian SV, Goldgar D, Borg A, Narod SA, Monteiro AN. 2005 Classification of BRCA1 missense variants of unknown clinical significance. J Med Genet. 42:138-146.

83. Situma C, Wang Y, Hupert M, Barany F, McCarley RL, Soper SA. 2005 Fabrication of DNA microarrays onto poly(methyl methacrylate) with ultraviolet patterning and microfluidics for the detection of low-abundant point mutations. Anal Biochem. 340:123-135.

84. Turner DJ, Zirvi MA, Barany F, Elenitsas R, Seykora J. 2005 Detection of the BRAF V600E mutation in melanocytic lesions using the ligase detection reaction. J Cutan Pathol. 32:334-339.

85. Hashimoto M, Hupert ML, Murphy MC, Soper SA, Cheng YW, Barany F. 2005 Ligase detection reaction/hybridization assays using three-dimensional microfluidic networks for the detection of low-abundant DNA point mutations. Anal Chem. 15:3243-3255.

86. Favis, R., Gerry, N.P., Cheng, Y.W. and Barany, F. (2005) Applications of the Universal DNA Microarray in Molecular Medicine. In "Methods in Molecular Medicine: Microarrays in Clinical Diagnostics" (Thomas O. Joos and Paolo Fortina, Editors), The Humana Press Inc., USA 114:25-58.

87. Soper SA, Hashimoto M, Situma C, Murphy MC, McCarley RL, Cheng YW, Barany F. (2005) Fabrication of DNA microarrays onto polymer substrates using UV modification protocols with integration into microfluidic platforms for the sensing of low-abundant DNA point mutations. Methods. 37:103-13.

88. Cheng YW, Shawber C, Notterman D, Paty P, Barany F. (2006) Multiplexed profiling of candidate genes for CpG island methylation status using a flexible PCR/LDR/Universal Array assay. Genome Res.16:282-289.

89. Hashinoto M, Barany F, Soper SA. (2006) Polymerase chain reaction/ligase detection reaction/hybridization assays using flow-through microfluidic devices for the detection of low-abundant DNA point mutations. Biosens Bioelectron. 21:2129-37.

90. Tsafrir D, Bacolod M, Selvanayagam Z, Tsafrir I, Shia J, Zeng Z, Liu H, Krier C, Stengel RF*, Barany F*, Gerald WL*, Paty PB*, Domany E*, Notterman DA*. (2006) Relationship of Gene expression and Chromosomal Abnormalities in Colorectal Cancer. Cancer Research 66:2129-37. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

91. Turner DJ, Pingle MR, Barany F. (2006) Harnessing asymmetrical substrate recognition by thermostable EndoV to achieve balanced linear amplification by thermophilic polymerase, and use in multiplexed SNP typing. Biochemistry and Cell Biology 84:232-42.

92. Wen Y, Giardina SF, Hamming D, Greenman J, Zachariah E, Bacolod MD, Liu H, Shia J, Amenta PS, Barany F*, Paty P*, Gerald W*, Notterman D*. (2006) GROalpha is highly expressed in adenocarcinoma of the colon and down-regulates fibulin-1. Clin Cancer Res.12:5951-9. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

93. Gao H, Huang J, Barany F, Cao W. (2007) Switching base preferences of mismatch cleavage in endonuclease V: an improved method for scanning point mutations. Nucleic Acids Res. 35(1):e2.

94. Pingle MR, Granger K, Feinberg P, Shatsky R, Sterling B, Rundell M, Spitzer E, Larone D, Golightly L, Barany F. (2007) Multiplexed identification of blood borne bacterial pathogens using a novel 16s rDNA PCR/LDR/Capillary Electrophoresis assay. J Clin Microbiol. 45: 1927-1935.

95. Gavert N, Sheffer M, Raveh S, Spaderna S, Shtutman M, Brabletz T, Barany F*, Paty P*, Notterman D*, Domany E*, Ben-Ze'ev A. (2007). Expression of L1-CAM and ADAM10 in human colon cancer cells induces metastasis. Cancer Res. 67:7703-12. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

96. Hashimoto M, Barany F, Xu F, Soper SA. (2007) Serial processing of biological reactions using flow-through microfluidic devices: coupled PCR/LDR for the detection of low-abundant DNA point mutations. Analyst. 132:913-21.

97. Forslund, A., Zeng, Z, Qin, L, Rosenberg, S., Ndubuisi, M., Pincas, H., Gerald, W*, Notterman D*, Barany F*, Paty P* (2008). Mdm2 gene amplification is correlated to tumor progression but not to presence of snp309 or TP53 mutational status in primary colorectal cancers. Molecular Cancer Research 6:205-211. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

98. Bacolod M, Schemmann G, Wang S, Shattock R, Giardina S, Zeng Z, Shia J, Stengel R, Gerry N, Hoh J, Kirchhoff T, Gold B, Christman M, Offit K, Gerald W*, Notterman D*, Ott J*, Paty P*, Barany F*. (2008). The Signatures of Autozygosity among Patients with Colorectal Cancer. Cancer Res. 68:2610-21. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

99. Khan SA, Idrees K, Forslund A, Zeng Z, Rosenberg S, Pincas H, Barany F*, Offit K, Laquaglia MP, Paty P*. (2008). Genetic variants in germline TP53 and MDM2 SNP309 are not associated with early onset colorectal cancer. J Surg Oncol. 97:621-5. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

100. Zeng ZS, Weiser MR, Kuntz E, Chen CT, Khan SA, Forslund A, Nash GM, Gimbel M, Yamaguchi Y, Culliford AT 4th, D'Alessio M, Barany F*, Paty P*. (2008) c-Met gene amplification is associated with advanced stage colorectal cancer and liver metastases. Cancer Lett. 265:258-69. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

101. Rondini S, Pingle MR, Das S, Tesh R, Rundell MS, Hom J, Stramer S, Turner K, Rossmann SN, Lanciotti R, Spier EG, Muñoz J, Larone D, Spitzer E, Barany F, Golightly LM. (2008) Development of a multiplex PCR/LDR assay for detection of West Nile Virus. J Clin Microbiol. 46(7):2269-79.

102. Das S, Pingle MR, Muñoz-Jordán J, Rundell MS, Rondini S, Granger K, Chang GJ, Kelly E, Spier EG, Larone D, Spitzer E, Barany F, Golightly LM. (2008) Detection and Serotyping of Dengue Virus in Serum Samples by Multiplex Reverse Transcriptase-PCR/LDR Assay. J Clin Microbiol. 46:3276-84.

103. Wang S, Haynes C, Barany F*, Ott J*. (2009) Genome-wide autozygosity mapping in human populations. Genet Epidemiol. 2009 Feb;33(2):172-80. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

104. Cheng YW, Pincas H, Bacolod MD, Schemmann G, Giardina SF, Huang J, Barral S, Idrees K, Khan SA, Zeng Z, Rosenberg S, Notterman DA*, Ott J*, Paty P*, Barany F*. (2008) CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. Clin Cancer Res. 2008 Oct 1;14(19):6005-13. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

105. Sinville R, Coyne J, Meagher RJ, Cheng YW, Barany F, Barron A, Soper SA. (2008) Ligase detection reaction for the analysis of point mutations using free-solution conjugate electrophoresis in a polymer microfluidic device. Electrophoresis 29(23):4751-4760.

106. Bacolod M, Schemmann G, Giardina S, Paty P*, Notterman D*, Barany F*. (2009) Emerging paradigms in cancer genetics: some important findings from high-density SNP array studies. Invited Review. Cancer Res. 2009 Feb 1;69(3):723-7. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

107. Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F*, Paty PB*, Gerald WL*, Notterman DA*, Domany E*. (2009) Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proc Natl Acad Sci U S A. 2009 Apr 28;106(17):7131-6. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

108. Nash GM, Gimbel M, Cohen AM, Zeng ZS, Ndubuisi MI, Nathanson DR, Ott J*, Barany F*, Paty PB*. (2009) KRAS Mutation and Microsatellite Instability: Two Genetic Markers of Early Tumor Development That Influence the Prognosis of Colorectal Cancer. Ann Surg Oncol. 2009 Oct 8. [Epub ahead of print] . *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

109. Granger K, Rundell MS, Pingle MR, Shatsky R, Larone DH, Golightly LM, Barany F, Spitzer ED. (2009) Multiplex PCR-Ligation Detection Reaction assay for the simultaneous detection of drug resistance and toxin genes from Staphylococcus aureus, Enterococcus faecalis and Enterococcus faecium. J Clin Microbiol. 2009 Oct 28. [Epub ahead of print].

110. Cheng YW, Idrees K, Shattock R, Khan SA, Zeng Z, Brennan CW, Paty P*, Barany F*. (2009) Loss of imprinting and marked gene elevation are two forms of aberrant IGF2 expression in colorectal cancer.Int J Cancer. 2009 Dec 2. [Epub ahead of print]. *Co-investigators of an international collaboration on molecular profiling of colon tumors; Program Director: Barany, F.

LECTURES

1. Plasmid exchange between *Streptococcus pneumoniae* and *E. coli.* Wind River Conference of Genetic Exchange, Estes Park, CO, June 9, 1981.

2. Plasmid exchange between Gram-positive and Gram-negative bacteria. Northwestern University Medical and Dental School, Chicago, IL, August 5, 1981.

3. Single-stranded DNA transformation and gene expression of fl- plasmid hybrids in *Streptococcus pneumoniae* and *Escherichia coli.* ASM International Conference on *Streptococcal* Genetics, Sarasota, FL, November 10, 1981.

4. *E. coli* insertion elements that increase expression of heterologous genes. New York Public Health Research Institute, New York, December 15, 1981.

5. Heterologous gene expression in Gram-positive and Gram-negative bacteria. University of Illinois at Chicago Circle, Chicago, IL. April 10, 1982.

6. How DNA eludes restriction during *Haemophilus* transformation. Pharmacia-P.L. Biochemicals, Inc., Milwaukee, WI, August 17, 1983.

7. The mechanism of DNA uptake and integration in *Haemophilus* transformation. Keynote speaker at Mid-Atlantic Extrachromosomal Genetics Elements Meeting, Virginia Beach, VA, October 1, 1983.

8. Directional transport and integration of donor DNA during *Haemophilus* transformation. Temple University School of Medicine, Philadelphia, PA, June 14, 1984.

9. Single-stranded hexameric oligonucleotide linkers for *in vitro* mutagenesis. Helen Hay Whitney Foundation meeting, Arden House, Harriman, NY, December 8, 1984.

10. Two-codon insertion mutagenesis of plasmid genes. University of California, Berkeley, Naval Biosciences Laboratory, Oakland, CA, March 7, 1985.

11. Single-stranded hexameric oligonucleotide linkers: Use for *in vitro* mutagenesis. Mutagenesis workshop UCLA Symposia - Protein Structure, Folding and Design. Keystone, CO, April 1, 1985.

12. Two-codon insertion mutagenesis. Invited speaker at 15th Linderstrom-Lang Conference of the Swedish Society for Microbiology and the Swedish Biochemical Society, Umea, Sweden, September 23, 1985.

13. Insertion mutagenesis of bacterial genes. NY Prokaryotic Molecular Biologists, The Rockefeller University, NY, May 27, 1986.

14. Two-codon insertion mutagenesis of the *Taq*I restriction endonuclease. New England Biolabs, Beverly, MA, December 4, 1986.

15. Analysis of functional domains using two-codon mutagenesis. Harvard Medical School, Boston, MA, December 5, 1986.

16. Analysis of protein functional domains using two codon insertion mutagenesis. University of Chicago, Chicago, Il. April 14, 1987.

17. Analysis of functional domains using two codon insertion mutagenesis. Invited speaker at Cornell University Biotechnology Symposium: Genetic engineering of proteins. Ithaca, N.Y. October 20, 1987.

18. A *Taq* * is born. New England Biolabs, Beverly, MA, December 10, 1987.

19. How *Taq*I restriction endonuclease recognizes its cognate sequence. University of Nebraska, Lincoln NE. February 17, 1988.

20. How *Taq*I recognizes its cognate sequence. Cold Spring Harbor Laboratories Cold Spring Harbor N.Y. May 17, 1988.

21. Sequence specific recognition of DNA by *Taq*I endonuclease. Workshop on Biological DNA Modification. Gloucester MA, May 21, 1988.

22. How *Taq*I restriction endonuclease recognizes its cognate sequence. Max-Planck-Institut fur Molekulare Genetik, Berlin West Germany, September 19, 1988.

23. How *Taq*I restriction endonuclease recognizes its cognate sequence. SIBIA, San Diego CA, November 15, 1988.

24. How *Taq*I restriction endonuclease recognizes its cognate sequence. University of California San Francisco, San Francisco CA, November 18, 1988.

25. How *Taq*I restriction endonuclease recognizes its cognate sequence. Abbott Laboratories, Chicago Il, November 21, 1988.

26. How *Taq*I restriction endonuclease recognizes its cognate sequence. Dupont, Central Res. and Dev. Dept. Wilmington, DE, January 25, 1989.

27. How *Taq*I restriction endonuclease recognizes its cognate sequence. Brookhaven National Labs. Upton, Long Island, NY, January 26, 1989.

28. How *Taq*I restriction endonuclease recognizes its cognate sequence. Hunter College, New York, NY, October 27, 1989.

29. How *Taq*I restriction endonuclease recognizes its cognate sequence. University of Rochester, Rochester, NY, November 30, 1989.

30. How *Taq*I restriction endonuclease recognizes its cognate sequence. California Institute of Technology, Pasadena, CA, January 25, 1990.

31. Detection of genetic diseases using thermostable DNA ligase. Genome Mapping and Sequencing Conference, Cold Spring Harbor, NY, May 3, 1990.

32. The exquisite specificity of *Thermus aquaticus* DNA recognition proteins. University of Maryland at Baltimore, Baltimore, MD, May 14, 1990.

33. The exquisite specificity of *Thermus aquaticus* DNA recognition proteins. Johns Hopkins Universtiy School of Medicine, Baltimore, MD, May 15, 1990.

34. The exquisite specificity of *Thermus aquaticus* DNA recognition proteins; Correlations between codon insertion mutants, revertants, and the three dimensional structure of ß-lactamase. International Symposium on Site-Directed Mutagenesis and Protein Engineering, (DNA/Protein interactions session Chairman). Tromso, Norway, August 29 & 30, 1990.

35. *TaqI* endonuclease insertion mutants and a comparison of its sequence with the *TthHB8I* isoschizomer. Invited speaker at Second New England Biolabs Workshop on Biological DNA Modification. West Berlin, Germany, September 3, 1990.

36. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Research Institute of Molecular Pathology, Vienna, Austria, September 7, 1990.

37. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. University College and Middlesex School of Medicine. London, England, September 10, 1990.

38. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Stratagene. San Diego, CA, November 16, 1990.

39. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Cetus. Emeryville, CA, November 19, 1990.

40. The exquisite specificity of *Thermus aquaticus* DNA recognition proteins. College of Physicians & Surgeons of Columbia University. New York, NY, January 18, 1991.

41. Single-nucleotide genetic disease detection using cloned thermostable ligase. Invited speaker at Miami/Biotechnology winter symposia. Advances in gene technology: The molecular biology of human genetic disease. Miami, FL, January 31, 1991.

42. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Mount Sinai Medical School. New York, NY, February 5, 1991.

43. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. University of Massachusetts Medical Center. Worcester, MA, February 22, 1991.

44. Detection of genetic diseases. 50th Westinghouse Science Talent Search Alumni Reunion. Washington DC. March 3, 1991.

45. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Public Health Research Institute, New York. April 2, 1991.

46. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Cornell University, Ithaca, NY, April 25, 1991.

47. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Centers for Disease Control, Atlanta, GA, May 14, 1991.

48. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Rocky Mountain Laboratory, Hamilton, MT, August 16, 1991.

49. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. BioRad, San Francisco, CA, November 25, 1991.

50. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Applied Biosystems Inc. Foster City, CA February 18, 1992.

51. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Max-Planck-Institut fur Molekulare Genetik, Berlin, Germany, March 23, 1992.

52. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Institut fur Mikrobiologie und Molekularbiologie, Giessen Germany, March 26, 1992.

53. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. Institut Pasteur, Paris, France, March 31, 1992.

54. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. University of Leicester, Leicester, England, April 3, 1992.

55. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. University of Bristol, Bristol, England, April 7, 1992.

56. *Thermus aquaticus* DNA recognition proteins, and their use for detection of genetic diseases. SUNY at Stonybrook, Stonybrook, NY, April 16, 1992.

57. Single-nucleotide disease detection using ligase chain reaction. Invited speaker at the 92nd American Society for Microbiology Meeting, New Orleans, LA, May 29, 1992.

58. The ligase chain reaction (LCR) for detection of mutations. Invited speaker at Seventh annual workshop on recent advances in molecular pathology, Tufts University School of Medicine, Boston, MA, June 19, 1992.

59. Thermophilic DNA recognition proteins, and their use for detection of genetic diseases. University of Illinois College of Medicine, Chicago, IL, November 5, 1992.

60. New concepts in cancer detection. Applied Biosystems Inc. Foster City, CA, March 25, 1993.

61. Genetic disease detection. Roche Molecular Systems. Alameda, CA, March 26, 1993.

62. Detection of genetic diseases. University of Illinois at Chicago, Chicago, IL, April 30, 1993.

63. Detection of genetic and infectious diseases using DNA diagnostics. Invited speaker at; PCR: Alternative technologies and applications, Boston, MA, June 7, 1993.

64. A biochemical analysis of the *Taql* restriction endonuclease. Invited speaker at; Restriction endonucleases and modification methyltransferases: Structures and mechanisms, FASEB research conference, Saxton River, VE, July 6, 1993.

65. Genetic disease detection. Strang Cancer Research Laboratory, New York, NY, December 13, 1993.

66. Genetic disease detection. Skirball Institute of Biomolecular Medicine, New York, NY, February 17, 1994.

67. New methods of detecting genetic diseases and cancers. Perkin Elmer/Applied Biosystems Foster City, CA, March 17, 1994.

68. New methods of detecting genetic diseases and cancers. Dean's Hour Lecture. Cornell University Medical College, New York, NY, March 23, 1994.

69. New methods of detecting genetic diseases and cancers. Yale University School of Medicine, New Haven, CT April 27, 1994.

70. New methods of detecting genetic diseases and cancers. University of Maryland, Baltimore, MD, May 2, 1994.

71. New methods of detecting genetic diseases and cancers. Wayne State University, Detroit, MI May 9, 1994.

72. New methods of detecting genetic diseases and cancers. Memorial Sloan Kettering Institute, New York, NY December 8, 1994.

73. New methods of detecting cancers. Oncor, Gaithersburg, MD, January 5, 1995.

74. New methods of detecting genetic diseases and cancers. Myriad, Salt Lake City, UT,February 6, 1995.

75. New methods of detecting genetic diseases and cancers. City of Hope Beckman Research Center, Duarte, CA, February 7, 1995.

76. New methods of detecting genetic diseases and cancers. Applied Biosystems Division of Perkin Elmer, Foster City, CA, February 8, 1995.

77. New methods of detecting genetic diseases and cancers. National Cancer Institute, Rockville, MD, March 8, 1995.

78. Detecting genetic diseases and cancers. National Institutes of Health, Rockville, MD, March 9, 1995.

79. New methods of detecting genetic diseases and cancers. Cold Spring Harbor Laboratories, Cold Spring Harbor, NY April 3, 1995.

80. New methods of detecting genetic diseases and cancers. Invited Speaker at "Accelerating Gene Discovery and Mutation Detection" conference, New York, NY, May 16, 1995.

81. New methods of detecting genetic diseases and cancers. Memorial Sloan Kettering Institute, New York, NY, May 31, 1995.

82. New methods of detecting genetic diseases and cancers. Wadsworth Center for Laboratories and Research, Albany, NY, June 6, 1995

83. New methods of detecting genetic diseases and cancers. Memorial Sloan Kettering Institute, New York, NY June 7, 1995.

84. New methods of detecting genetic diseases and cancers. National Cancer Institute, Early Detection Research Network Site Visit, Rockville, MD, June 8, 1995.

85. New approaches to cancer detection. Strang Cancer Prevention Center Annual Meeting, New York, NY, June 9, 1995.

86. New methods of detecting genetic diseases and cancers. Invited speaker at "Applications of Diagnostics in Health Care, the Environment, and Agriculture." Cornell University, Ithaca, NY, October 9, 1995.

87. New methods of detecting genetic diseases and cancers. Yale University Medical School,New Haven, CT, November 21, 1995.

88. New methods of detecting genetic diseases and cancers. Chiron, Emmeryville, CA, December 5, 1995.

89. Vistas in genetic analysis. University of Illinois at Chicago. "Biochemistry and Pathophysiology of Muscle. A symposium in tribute to Professors Michael and Kate Barany." Chicago, Il, May 13, 1996.

90. New methods of detecting genetic diseases and cancers. Motorola, Phoenix, Az, May 20, 1996.

91. DNA recognition proteins and their use in detecting genetic diseases and cancers. Invited speaker at "Enzymes that act on Nucleic Acids" FASEB research conference, Saxton River, VE, June 20, 1996.

92. Development of programable DNA arrays. Progress report on the ATP/ NIST joint project. Applied Biosystems Division of Perkin Elmer, Foster City, CA, August 15, 1996.

93. (i) Improving the fidelity of thermostable DNA ligase. (ii) Multiplexed detection of K-*ras* mutations in colorectal cancer. Applied Biosystems Division of Perkin Elmer, Foster City, CA, August 16, 1996.

94. New methods of detecting genetic diseases and cancers. Johnson & Johnson Diagnostics, Rochester, NY, August 26, 1996.

95. New methods of detecting genetic diseases and cancers. Merck Research Laboratories, West Point, PA, October 10, 1996.

96. Early detection of colon cancer mutations. New York Human Genetics Club, Columbia University College of Physicians & Surgeons, NY, October 17, 1996.

97. DNA recognition proteins and their use in detecting genetic diseases and cancers. The Institute of Genomic Research, Rockville, MD, October 18, 1996.

98. New methods of detecting genetic diseases and cancers. City of Hope Beckman Research Center, Duarte, CA, October 23, 1996.

99. Multiplexed detection of genetic and forensic polymorphisms. Motorola, Schaumburg, IL , November 7, 1996

100. New methods of detecting genetic diseases and cancers. Perkin Elmer Wilton, CT November 18, 1996.

101. New methods of detecting genetic diseases and cancers. Novel Amplification Technologies (Chairman: Update on the latest Amplification Technologies), Washington, D.C. December 16, 1996.

102. New approaches to detecting repeat sequence polymorphisms associated with colorectal cancer. Motorola, Schaumburg, IL , April 11, 1997.

103. New approaches to detecting point mutations and repeat sequence polymorphisms associated with spontaneous colorectal cancer. Applied Biosystems Division of Perkin Elmer, Foster City, CA, April 24, 1997.

104. Fidelity and Error in Thermophilic DNA-Recognition Proteins. The Rockefeller University, New York, NY, January 13, 1998.

105. New Methods of Cancer Detection. Yale University Medical School, New Haven, CT, January 27, 1998.

106. Thermophilic DNA recognition proteins, and their use for detection of cancer-associated mutations. . Co-organizer and speaker at "Nucleic Acid Enzymes: Mechanisms and Diseases " FASEB research conference,   Saxton River, VE, June 16, 1998.

107. New Methods of Cancer Detection. Abbott Labs, Chicago, Il, October 22, 1998.

108. New Methods of Cancer Detection. Columbia University College of Physicians & Surgeons, NY, October 26, 1998.

109. New Methods of Cancer Detection. Mount Sinai Medical School, NY, October 27, 1998.

110. New Methods of Cancer Detection. Public Health Research Institute, NY, November 3, 1998.

111. Around the Genome in 80 days. Celera, Rockville, MD. January 7, 1999.

112. New Methods of Cancer Detection. National Cancer Institute. Bethesda, MD. March 25, 1999.

113. New Methods of Cancer Detection. Institute of Biotechnology, San Antonio, Tx. April 6, 1999.

114. New Methods of Cancer Detection. University of Texas Austin, Austin, Tx. April 7, 1999.

115. Cancer Detection from Clinical Samples: Future Challenges for Nanofabricated Devices. Invited speaker at Nanofabricated Devices Conference, San Jose, CA, April 20, 1999.

116. Universal DNA arrays and new enzymes for cancer detection. PE-Biosystems, Foster City, CA, April 21, 1999.

117. New Methods of Cancer Detection. National Cancer Institute. Bethesda, MD. April 28, 1999.

118. New Methods of Cancer Detection. Institute Pasteur. Paris, France. May 31, 1999.

119. New Methods of Cancer Detection. Karolinska Institute. Stockholm, Sweden. June 1, 1999.

120. New Methods of Cancer Detection. Institute Curie. Paris, France. June 2, 1999.

121. Multiplex detection of cancer mutations using ligase based assays: application to p53, K-ras, met oncogene, APC, BRCA1, and BRCA2. Invited speaker and session chair at Eurocancer 99, Paris, France, June 3, 1999.

122. Use of DNA recognition proteins to identify genome changes in cancers. Invited speaker at "Nucleic Acid Enzymes: Structures, Mechanisms and Novel Applications" FASEB research conference, Saxton River, VE, June 21, 2000.

123. New Methods of Cancer Detection. Institut de Genetique Moleculaire. Montpellier, France. August 28, 2000.

124. New Methods of Cancer Detection. Institute Curie. Paris, France. August 29, 2000.

125. New Methods of Cancer Detection. Invited Keynote Speaker at "Arrays and Beyond" Wadsworth Center, Albany, NY, December 5, 2000.

126. New Methods of Cancer Detection. Distinguished Speakers Seminar Program, National Cancer Institute, Frederick Cancer Research and Development Center. Ft. Detrick, MD. December 12, 2000.

127. New Methods of Cancer Detection. Invited Speaker at American Society for Clinical Oncologists (ASCO). San Francisco, CA. May 12 and 14, 2001.

128. New Methods of Cancer Detection. Invited Speaker at Chips-to-Hits IBC meeting. San Diego, CA. November 1, 2001.

129. Molecular profiling of colon tumors. Applied Biosystems, Foster City, CA, February 19, 2002.

130. New Methods of Cancer Detection. University of South California, Los Angeles, CA, February 21, 2002.

131. New Methods of Cancer Detection. Purdue University, West Lafayette, IN March 18, 2002.

132. New Methods of Cancer Detection. Invited Speaker at Lennox K. Black Symposium – "Genomics & Bioinformatics For The Advancement Of Clinical Science" Philadelphia, PA, Oct. 13, 2002.

133. New Methods of Cancer Detection. Invited Speaker at American Association of Cancer Researhers (AACR). Frontiers in Cancer: Prevention Research, Boston, MA Oct. 14, 2002.

134. New Methods of Cancer Detection. MIT, Boston, MA Oct. 17, 2002.

135. Potential Benefits of Molecular Profiling to Cancer Diagnosis. Plenary Address, Invited Speaker, at Chips-to-Hits IBC meeting. Philadelphia, PA. October 30, 2002.

136. Molecular Profiling of Tumors. Grand Rounds, Dept. of Pathology, Weill Medical College of Cornell University, New York, NY, November 25, 2002.

137. Harmonized Microarray / Mutation Scanning Analysis of Colorectal Tumors. Invited Speaker and Session Chair. 7th Mutation Detection Workshop, Palm Cove, Queensland, Australia July 3, 2003.

138. Harmonized Microarray / Mutation Scanning and Methylation Analysis of Colorectal Tumors. Keynote Address. BioArrays-2003-New York, New York, NY October 1, 2003.

139. Multiplexed pathogen detection for Biodefense. Applied Biosystems, Foster City, CA, February 18, 2004.

140. Molecular profiling of colon tumors. Louisiana State University. Baton Rouge, LA May 13, 2004

141. Molecular profiling of Cancer. Invited Keynote Speaker. Pharmocogenomics / Toxicogenomics Johnson & Johnson Symposium, New Brunswick, NJ, September 29, 2004.

142. Molecular profiling of colon tumors. Applied Biosystems, Foster City, CA, October 25, 2004.

143. Molecular profiling of tumors. Keynote Address at Clinical Genomics IBC conference, San Diego, CA, February 17, 2005.

144. Molecular profiling of colon tumors. National Cancer Institute, Washington, DC, March 16, 2005.

145. Molecular profiling of tumors. Invited Speaker at Technology Fair 2005. United States Patent Office, Washington, DC, March 17, 2005.

146. Molecular Detection and Diagnosis: The role of detection and rapid diagnosis in treating infectious disease. Invited Speaker at National Academy of Sciences Workshop on "New directions in the study of antimicrobial therapeutics: New classes of antimicrobials". Washington, DC, March 24, 2005.

147. Molecular profiling of cancers. Weizmann Institue of Science, Rehovot, Israel, September 4, 2005.

148. Molecular techniques for tumor investigations. Invited speaker. British Human Genetics Conference, York, England, September 14, 2005.

149. Molecular profiling of cancers. University of Notre Dame, Notre Dame, IN, October 10, 2005.

150. Multiplexed detection of biothreat agents. New York Department of Public Health, NY, January 27, 2006.

151. Molecular profiling of cancers. Celera Diagnostics. Alameda, CA, February 21, 2006.

152. Molecular profiling of cancers. Roche Molecular Systems. Alameda, CA, February 21, 2006.

153. Multiplexed detection of blood-borne pathogens. Cepheid. Sunnyvale, CA, February 22, 2006.

154. Molecular profiling of colon tumors. Affymetrix. Santa Clara, CA, February 22, 2006.

155. Multiplexed Blood-Borne Pathogen Identification and Detection. National Institute of Allergy and Infectious Diseases, Washington, DC, August 17, 2006.

156. Molecular profiling of colon tumors. Invited Speaker at Chips-to-Hits IBC meeting. Boston, MA. September 27, 2006.

157. Molecular profiling of colon tumors. Grand Rounds, Dept. of Pathology, Weill Medical College of Cornell University, New York, NY October 9, 2006.

158. Molecular profiling of colon tumors. Grand Rounds, Dept. of Pathology, College of Physicians & Surgeons of Columbia University. New York, NY, October 17, 2006.

159. Molecular profiling of colon tumors. Invited Speaker: Colon Cancer Initiative Meeting. The Ludwig Institute for Cancer Research – Hospital A. Oswaldo Cruz, Sao Paolo, Brazil, February 21, 2007.

160. Molecular profiling of colon tumors. Cepheid. Sunnyvale, CA, June 25, 2007.

161. Molecular profiling of cancers. Ensemble, Boston, Sept. 17, 2007.

162. Molecular profiling of colon tumors. Invited Speaker at AACR Special Conference: Advances in Colon Cancer Research meeting. Boston, MA. November 17, 2007.

163. Molecular profiling of Colon Tumors in 4 dimensions. Colon Cancer Initiative Meeting. The Ludwig Institute for Cancer Research, New York, NY, November 28, 2007.

164. Molecular profiling of Colon Tumors: From gene discovery to individualized therapy. Genentech, South San Francisco, CA, January 11, 2008.

165. Molecular profiling of colon tumors. Moffitt Cancer Center, Tampa,Fl February 21, 2008.

166. Molecular profiling of colon tumors. MD Anderson Cancer Center, Houston, Tx March 18, 2008.

167. Coferon drug design platform. Invited Speaker at Cornell Center for Technology Enterprise and Commercialization (CCTEC) Emerging Technologies Showcase - New Technologies, Cornell University, Ithaca, NY April 10, 2008.

168. Coferons: A new paradigm for drug design. Ensemble and Flagship Enterprises, Boston, MA, April 17, 2008.

169. Coferons: A new paradigm for drug design. Genentech, South San Francisco, CA, April 21, 2008.

170. Molecular profiling of Cancers and Infectious Diseases. Cepheid. Sunnyvale, CA, April 22, 2008.

171. Molecular profiling of colon tumors. Genome Institute of Singapore, Singapore April 30, 2008.

172. Identification of colon cancer genes inherited by descent. Invited speaker at LICR Colon Cancer Initiative, Melbourne, Australia, May 5, 2008.

173. Molecular profiling of colon tumors. Purdue University, West Lafayette, IN November 17, 2008.

174. Molecular profiling of colon tumors. Indiana University School of Medicine, Indianapolis, IN November 18, 2008.

175. Rx-Gen Sequencing for Personalized Cancer Care. Presentation, and Workshop organized by F. Barany and JP. Robinson, at the invitation of Roche, Sony, & Parker Hannifin, San Francisco, CA, February 18, 2009.

176. Rx-Gen Sequencing for Personalized Cancer Care. Presentation, and Workshop organized by F. Barany and JP. Robinson, at the invitation of Roche & Roche 454, Bramford CT, March 16, 2009.

177. Rx-Gen Sequencing for Personalized Cancer Care. Roche Molecular Diagnostics, San Francisco, CA, April 29, 2009.

February 19, 2010

Exhibit 2:   Naiser et al., "Impact of Point-Mutations on the Hybridization Affinity of
              Surface-Bound DNA/DNA and RNA/DNA Oligonucleotide-Duplexes:
              Comparison of Single Base Mismatches and Base Bulges," *BMC Biotech.* 8:48
              (2008)

BMC Biotechnology

BioMed Central

Research article

Open Access

Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: Comparison of single base mismatches and base bulges

Thomas Naiser*[1], Oliver Ehler[1], Jona Kayser[1], Timo Mai[1], Wolfgang Michel[1] and Albrecht Ott[1,2]

Address: [1]Experimentalphysik I, Universität Bayreuth, D-95440 Bayreuth, Germany and [2]Experimentalphysik, Universität des Saarlandes, D-66041 Saarbrücken, Germany

Email: Thomas Naiser* - thomas.naiser@ep1.uni-bayreuth.de; Oliver Ehler - oliver.ehler@ep1.uni-bayreuth.de; Jona Kayser - jona.kayser@ep1.uni-bayreuth.de; Timo Mai - timo.mai@ep1.uni-bayreuth.de; Wolfgang Michel - wolfgang.michel@ep1.uni-bayreuth.de; Albrecht Ott - albrecht.ott@physik.uni-saarland.de

* Corresponding author

Abstract

Background: The high binding specificity of short 10 to 30 mer oligonucleotide probes enables single base mismatch (MM) discrimination and thus provides the basis for genotyping and resequencing microarray applications. Recent experiments indicate that the underlying principles governing DNA microarray hybridization – and in particular MM discrimination – are not completely understood. Microarrays usually address complex mixtures of DNA targets. In order to reduce the level of complexity and to study the problem of surface-based hybridization with point defects in more detail, we performed array based hybridization experiments in well controlled and simple situations.

Results: We performed microarray hybridization experiments with short 16 to 40 mer target and probe lengths (in situations without competitive hybridization) in order to systematically investigate the impact of point-mutations – varying defect type and position – on the oligonucleotide duplex binding affinity. The influence of single base bulges and single base MMs depends predominantly on position – it is largest in the middle of the strand. The position-dependent influence of base bulges is very similar to that of single base MMs, however certain bulges give rise to an unexpectedly high binding affinity. Besides the defect (MM or bulge) type, which is the second contribution in importance to hybridization affinity, there is also a sequence dependence, which extends beyond the defect next-neighbor and which is difficult to quantify. Direct comparison between binding affinities of DNA/DNA and RNA/DNA duplexes shows, that RNA/DNA purine-purine MMs are more discriminating than corresponding DNA/DNA MMs. In DNA/DNA MM discrimination the affected base pair (C·G vs. A·T) is the pertinent parameter. We attribute these differences to the different structures of the duplexes (A vs. B form).

Conclusion: We have shown that DNA microarrays can resolve even subtle changes in hybridization affinity for simple target mixtures. We have further shown that the impact of point defects on oligonucleotide stability can be broken down to a hierarchy of effects. In order to explain our observations we propose DNA molecular dynamics – in form of zipping of the oligonucleotide duplex – to play an important role.

Background

DNA microarray technology relies on the highly specific binding affinity of surface-tethered DNA probe sequences to complementary target sequences. Nucleic acid hybridization, the sequential base pairing between complementary probe and target strands, results in the formation of stable double-stranded duplexes. In microarray hybridization assays single-stranded nucleic acid targets – contained in a complex mixture of different target sequences in solution – freely diffuse over the surface-tethered probes until they are captured by a complementary probe. Target strands often carry fluorescent dye labels to enable quantitative detection of the individual target species. Hybridized targets can be identified by the position of the corresponding microarray features (each containing one particular species of surface-tethered probe strands) within the regular grid of the DNA microarray.

In DNA microarray applications, along with a high binding affinity (providing sensitivity), a high specificity of probe-target hybridization is required to discriminate between sometimes very similar homologous sequences. Binding specificity is particularly important in genotyping applications where Single Nucleotide Polymorphisms (SNPs), genetic variations of single bases, are concerned. SNPs determine genetic individuality, but also predisposition to a variety of genetic diseases, response to drugs, pathogens, chemicals and other agents. SNPs are of great interest not only for genetic research but also for medical diagnostics and therapy [1,2]

SNPs and point-mutations can be detected by means of relatively short 10 to 30 mer probes: Already a single mismatched (MM) base pair can result in a significant decrease of the duplex binding affinity with respect to the corresponding perfect matching (PM) duplex [3]

The binding affinity of mismatched duplexes – in bulk solution – is commonly predicted on the basis of the nearest-neighbor model [4-6] A recent study by Pozhitkov et al. [7] revealed a poor correlation between predicted duplex binding affinities and actual hybridization signal intensities implying that the thermodynamic properties of oligonucleotide hybridization on DNA microarrays are by far not understood. In DNA microarray experiments the binding affinity of mismatched oligonucleotide duplexes is governed not just by nearest-neighbor parameters – as in solution-phase hybridization – but mainly by the position of the defect [7-10]. Furthermore, the secondary structure of the long target strands [11] and various surface effects [12] have a significant influence on the microarray binding affinity

Our study is a comprehensive approach to understand how point defects affect the hybridization of fluorescently labeled oligonucleotide targets to surface-bound oligonucleotide probes. Rather than previous work on single base MMs, which has been conducted with complex target mixtures either from PCR products [9] or in vitro transcripts [7], we employ short (20-37 nt), end-labeled oligonucleotide targets, thus avoiding labeling and steric hindrance related effects. In order to avoid competitive binding [13] we perform each hybridization assay with a single target sequence. Oligonucleotide target sequences (DNA and RNA – see Tab. 1) were chosen to minimize secondary structures and any related influence on the hybridization signal. In particular we investigated differences between the impact of defects on DNA/DNA and analogue RNA/DNA duplexes.

DNA chips were fabricated by light-directed in situ synthesis [14,15] with a digital micromirror device (DMD™, Texas Instruments) based maskless synthesis apparatus [16-21] developed in our laboratory [10]. Sets of probe sequences were derived from probe sequence motifs by

Table 1: Fluorescently labeled target oligonucleotides used in this study.

Name		Target sequence (5'→3')	Label	Length (nt)
URA	DNA	ACTACAAACTTAGAGTGCAG... ...CAGAGGGGAGTGGAATTC	5'-Cy3	38
NIE	DNA	ACTCGCAAGCACCACCCTATCA	3'-Cy3	22
LBE	DNA	GTGATGCTTGTATGGAGGAA... ...TACTGCGATT	3'-Cy3	30
PET	DNA	ACATCAGTGCCTGTGTACTAGGAC	3'-Cy3	24
BEI	DNA	ACGGAACTGAAAGCAAAGAC	3'-Cy3	20
COM	DNA	AACTCGCTATAATGACCTGGACTG	5'-Cy3	24
NCO	DNA	TAGTGGGAGTTGTTAGTGATGTGA	3'-Cy3	24
PET	RNA	ACAUCAGUGCCUGUGUACUAGGACA	5'-Cy3	25
LBE	RNA	GUGAUGCUUGUAUGGAGGAA .UACUGCGAUUCGAU	5'-Cy3	34
COM	RNA	AACUCGCUAUAAUGACCUGGACUG	5'-Cy3	24

Fluorescently labeled DNA and RNA target oligonucleotides.

systematic variation of defect type and defect position including all single base mismatches, insertions and deletions. The design of the hybridization experiments (Fig. 1) enables discrimination between the strong influence of defect position [7,9,10] and the more subtle defect-type and sequence related factors.

After the current article was submitted, we became aware of further related studies in this area. Suzuki *et al.* [22] performed hybridization on custom NimbleExpress™ arrays (Affymetrix Inc) to investigate the influence of the probe length and mismatch position on single base MM discrimination Fish *et al* [23] performed a direct comparison between hybridization signals (perfectly matching and mismatched duplexes) from spotted microarrays and measured thermodynamic melting parameters (determined by differential scanning calorimetry in bulk solution) They report a linear relation between the duplex free energy and the microarray hybridization intensity.

The focus of the present paper is on the impact of various defect types (single base mismatches and single base bulges) on the hybridization signal.

Results

Our microarray hybridization experiments performed in this study provide quantitative information on the binding affinity of individual mismatched duplexes by means of the hybridization signal intensity (fluorescence of hybridized targets) Since the absolute hybridization signal intensities of the different sequence motifs employed in this study (Tab. I) are subject to a large variation (often larger than between mismatched and corresponding PM hybridization signals) we compare the MM hybridization signals with the corresponding PM hybridization signals. Our experiments – experimental details (probe sets, hybridization signal normalization etc) are explained in the Methods section – provide a measure for the mismatch discrimination with respect to the corresponding PM binding affinity, rather than an absolute measure for the MM binding affinity The discrimination between the hybridization affinity of point-mutated probes and corresponding perfect matching probes depends on the stability of the particular probe sequence In agreement with [22] we observed that the (more stable) 25 mer probes are less discriminative with respect to point defects than the shorter 16 mer probes Discrimination is also reduced for sequence motifs stabilized by a higher CG content.

MM defect position and hybridization affinity

The "defect profile" plots (plots of the normalized hybridization signal vs. defect position - e g in Fig. 2) show that the dominant parameter determining oligonucleotide probe-target-affinity - on the microarray surface - is the position of the defect A moving average evidences a

trough-like "mean profile" curve (*solid black line* in Fig. 2) A parabolic fit can provide a reasonable approximation for the average position dependence obtained from a large number of different sequence motifs [7,9]. For 16 mer duplexes a single base mismatch in the center typically results in 40% of the PM hybridization signal However, for individual sequence motifs we found sequence-dependent deviations from the simple position dependence (see Fig 3). The raw signal intensities and probe/target sequences of the experiment are given in Additional file 1

### Influence of the mismatch type in DNA/DNA duplexes

In the following we use the notation of the mismatch base pair $X \cdot Y$ consisting of the mismatched base X in the probe sequence and the base Y in the target sequence To investigate how the particular MM-types $X \cdot Y$ affect duplex stability we measured probe-target-affinities for 25 different sequence motifs Microarray hybridization experiments with single base mismatch probe sets as well as the extraction of their hybridization signals, which reflect duplex stability, are described in more detail in the Methods section. Owing to the limited number of available target oligonucleotides we restricted base substitutions to the probe sequences. The PM hybridization signals of the different 16 mer sequence motifs display a strong variation (up to a factor 20). Since the relative hybridization signal intensities within the individual probe sets are largely unaffected by this variation, we normalize the "defect profiles" by division with their standard deviation. The resulting database comprising normalized hybridization signals from about 1000 different single MM probe sequences, enables categorization of the binding affinities according to the mismatch type.

For statistical analysis of MM type and nearest-neighbor influences the superposed positional influence needs to be eliminated by subtraction of the mean profile. The resulting position-independent defect profile (for simplicity we keep the term "defect profile") consisting of influences of defect type and defect neighborhood only is shown in Fig. 2B. The boxplot representation of this data in Fig. 4 demonstrates that MM-types affecting C·G base pairs (i.e. A·C, C·C, T·C and A·G, G·G, T·G) systematically have lower median hybridization signal values than MM-types affecting A·T base pairs (A·A, C·A, G·A and C·T, G·T, T·T).

We compared the MM-type related hybridization signal deviations δI_{mp} from the mean MM profiles (Fig 2B) to predicted Gibbs free energy differences $\delta \Delta G_{37}^{\circ} = \Delta G_{37MM}^{\circ} - \Delta G_{37PM}^{\circ}$ between MM and corresponding PM duplexes. $\delta \Delta G_{37}^{\circ}$ were determined from

5 - AACTCGCTATAATGACCTGGACTG-Cy3 - 3   Target oligonucleotide
        3'-TATTACTGGACCTGAC-5'   Probe sequence motif
                                 (complementary to a
                                 section of the target)

**Set of point-mutated probe sequences,
derived from common probe sequence motif**

3'-AATTACTGGACCTGAC-5' ⎫
3'-CATTACTGGACCTGAC-5' ⎬ Single base mismatch (MM) probes
3'-GATTACTGGACCTGAC-5' ⎭
3'-TATTACTGGACCTGAC-5'   Perfect match (PM) probe

Defect position 1

3'-TAATTACTGGACCTGAC-5' ⎫
3'-TCATTACTGGACCTGAC-5' ⎬ Single base insertion probes
3'-TGATTACTGGACCTGAC-5' ⎭
3'-TTATTACTGGACCTGAC-5'
3'- ATTACTGGACCTGAC-5'   Single base deletion probe

3'-TATTACTGGACCTGAC-5'
3'-TCTTACTGGACCTGAC-5'
3'-TGTTACTGGACCTGAC-5'
3'-TTTTACTGGACCTGAC-5'

Defect position 2

3'-TAATTACTGGACCTGAC-5'
3'-TACTTACTGGACCTGAC-5'
3'-TAGTTACTGGACCTGAC-5'
3'-TATTTACTGGACCTGAC-5'
3'-T TTACTGGACCTGAC-5'

**Feature arrangement on
the microarray**

Defect position 1

Defect position 2

Hybridization signals

**Hybridization with the
Target sequence**

Defect profile

Hybridization
signal (a.u.)

Defect position

Figure 1
Design of the experiment. A comprehensive set of point-mutated probes is derived from a common *probe sequence motif*
which is complementary to the *target* sequence. Probe sequences are shown for the first two defect positions only To enhance
quantitative analysis probe sequences are arranged on the microarray as a compact *feature block* Hybridization signals from
hybridization with the target sequence are plotted versus defect position The *defect profile* shows relative hybridization affini-
ties depending on the probe sequence motif, defect type and defect position.
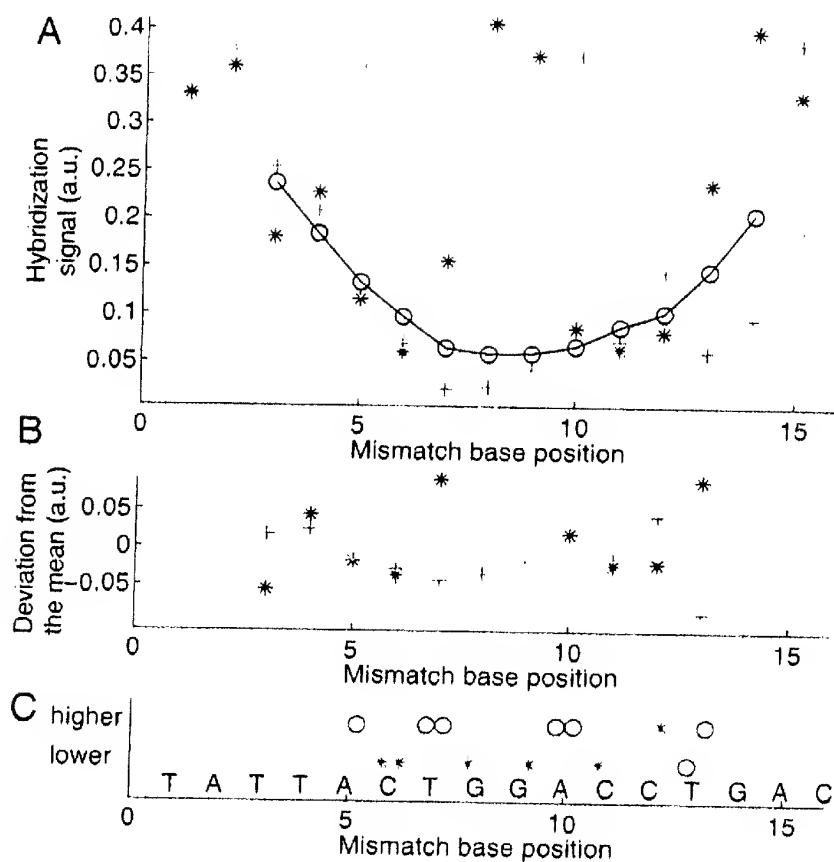
Figure 2
Mismatch defect profile (A) (hybridization signal vs. defect base position) obtained from the hybridization signals of the feature block shown in the right part of Additional file 9. Solution-background correction (see Methods section) was applied on raw hybridization signal intensities. The probe sequence motif 3'-TATTACTGGACCTGAC-5' is complementary to the target oligonucleotide COM. Markers depict the substituent base type (A red crosses; C green circles; G blue stars; T cyan triangles). The black line indicates the 'mean profile' (moving average of all mismatch hybridization signals over positions $p - 2$ to $p + 2$). PM probes, included as control to detect erroneous bias, have the largest hybridization signals (at a level of about 0.38 a.u.). The variation of the PM probe intensities also provides an estimate for the error of the measurement. Errors between distant microarray features, due to gradient effects, are expected to be larger than errors between the compactly arranged features corresponding to a particular defect position. **(B)** Deviation profile. The strong position dependent component of the hybridization signal is eliminated by subtraction of the mean profile. **(C)** Comparison of mean mismatch hybridization signals (average of the three mismatch hybridization signals at a particular defect position) at the sites of C·G base pairs to mean MM hybridization signals at the site of adjacent A·T base pairs. A marker (red star: A·T; blue circle C·G) is set in the upper row if the hybridization signal of the mismatches at the corresponding site is higher than at the adjacent site; otherwise a marker is set in the lower row. We noticed that mismatches substituting a C·G base pair usually have systematically lower hybridization signals than mismatches substituting a neighboring A·T base pair
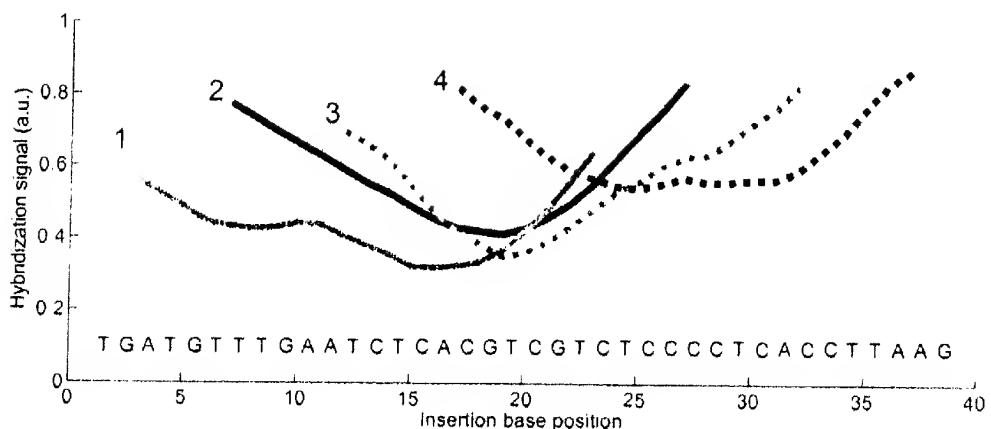
Figure 3
The impact of defects is affected by the local sequence environment. Single base insertion profiles (hybridization signal plotted versus the insertion base position) of four 25 mer probe sequence motifs complementary to the same target URA. Following solution-background correction of the raw intensity data (Methods section) hybridization signals were normalized with respect to the largest hybridization signal in each of the four insertion profiles. The probe motifs 1 to 4 hybridize at different sections of the target oligonucleotide. Mean profiles (thick lines) were obtained from the moving average of the particular insertion profiles (particular hybridization signal are shown as faint symbols – profile 4 is shown in detail in Fig. 5A). The mean profiles 1 to 3 have a distinct minimum between base positions 15 to 20. The stabilizing CG-rich region following after base position 20 results in increased hybridization signals in profile 4.

mismatch nearest-neighbor thermodynamic parameters [4]. Our analysis (shown in Additional file 2) indicates a decreasing trend of the $\mathcal{B}_{mp}$ values with increasing $\delta\Delta G^{1}_{17}$. Moreover, we observed that single base mismatches with two A·T flanking base pairs tend to provide a better mismatch discrimination than mismatches flanked by two C·G base pairs

DNA/DNA single base bulge defects
Single base insertions and deletions owing to a surplus unpaired base in one of the two strands result in bulged duplexes. In our experiments (sequence data and hybridization signal raw data is provided in Additional file 3) the bulged base is located on the surface-bound probe strand, whereas in duplexes with single base deletions (on the probe sequence) the bulge is on the target strand.

We discovered that on average the positional dependence of the *insertion* and *deletion defect profiles* (e.g. in Figs. 3 and 5A) is very similar to the positional dependence of *mismatch defect profiles* (Fig 2). Within one and the same individual defect profile, single base bulge defects originating from single base insertions or deletions display the same positional dependence as single base mismatch defects

(direct comparison shown in Fig. 6 – hybridization signal data provided in Additional file 4), qualitatively as well as quantitatively. On average single base insertion probes provide increased hybridization signals when compared to MM probes or single base deletions (Fig. 7). Besides the significantly increased hybridization signals of *Group II* insertions (see below), this is due to the reduced number of binding base pairs in the mismatched duplexes (which have one binding base pair less than the PM duplex, whereas a single base insertion leaves the number of binding base pairs unchanged). In single base insertions no binding base pair is substituted, but we see that the influence of the inserted base clearly depends on its neighbor The individual curves (e.g. the curve of C insertions - green circles in Fig. 5) show deviations from the (moving average) mean profile, hybridization signals can be significantly increased over several consecutive defect positions In particular base insertions next to identical bases (so called *Group II bulges* [24]) result in systematically increased binding affinities – in comparison to insertions of non-identical bases (*Group I bulges*) *Group II* bulges located near the center of 16 mer probes often show hybridization signals with a similar intensity as the corresponding PM probe (Fig 6, Fig 5C). A statistical analysis with a large dataset (Fig 8) comprising hybridization sig-
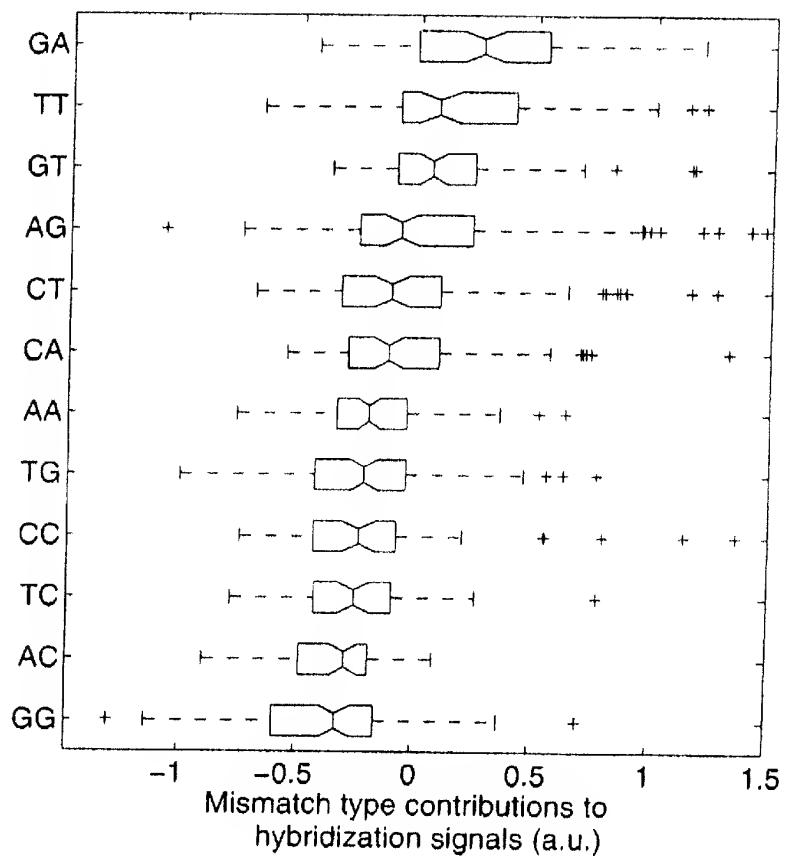
Figure 4
Boxplot representation of the hybridization signal distributions for the individual mismatch types, arranged according to the median values (the 95% confidence bounds are depicted by the notch). Boxes indicate the inter-quartile range (from the 25th to 75th percentile) containing 50% of the data. Whiskers extend to a maximum value of 1.5 times the interquartile range from the boxes ends. differ significantly with a 95 percent confidence Data processing: raw intensity data, solution-background correction, subtraction of the mean profile, normalization of the defect-type dependent deviations from the mean profile by division by the standard deviation of the defect profile (see Methods section). The mismatch types with the lowest hybridization signals are those (T·G, C·C, T·C, A·C, G·G) where C·G base pairs are affected by the mismatch defect. The only exception is A·G. The positive tails of this and other distributions seem to originate from stabilizing C·G base pairs next to the defect. ΔG°_{37} (standard deviation assuming that the various MM nearest-neighbor types are equally distributed)
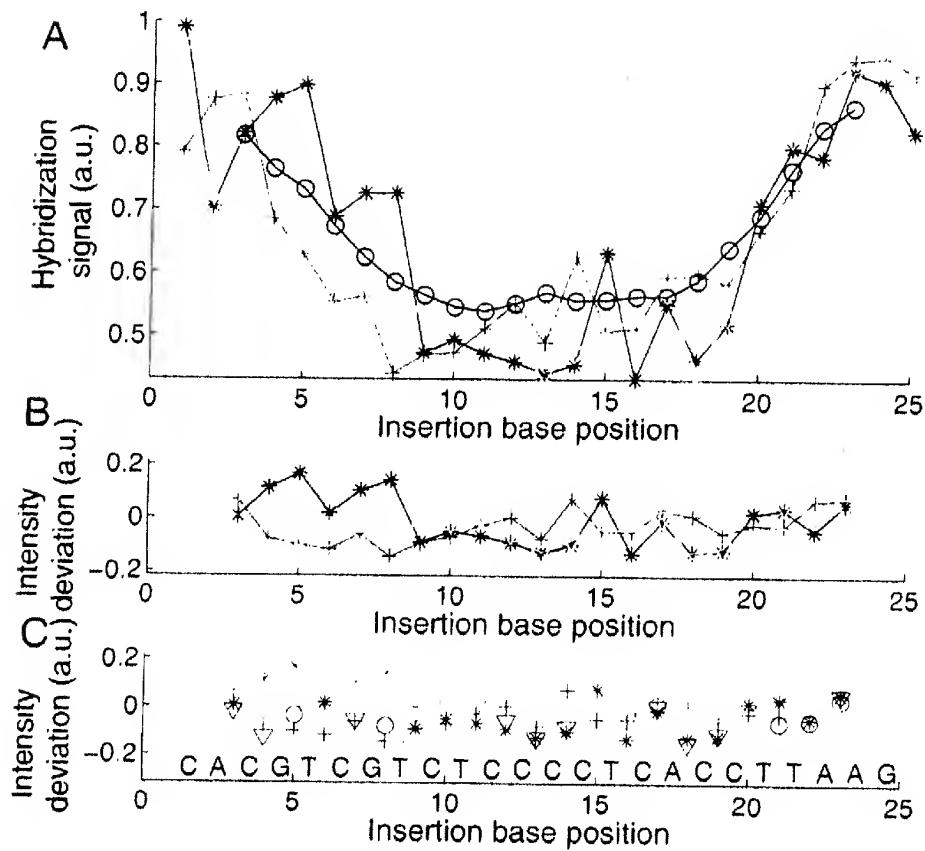
Figure 5
(A) Single base Insertion defect profile (hybridization signal plotted against the insertion base position; follow-ing solution background-correction of the raw intensity data, hybridization signals were normalized with respect to the largest hybridization signal in the insertion profile) of the probe sequence motif 3'-CACGTCGTCTCCCCTCACCTTAAG-5' (complementary to the target URA). Symbols correspond to insertion bases (A red crosses; C green circles; G blue stars; T cyan triangles). The mean profile (black line), obtained from the moving average (including all 4 insertion types) over positions $p - 2$ to $p + 2$ shows the common positional dependence. Insertions to the left and to the right of an identical base (*Group II bulges* – see text) result in identical probe sequences. **(B)** and **(C)** Devia-tion profiles. Positional influence is mostly eliminated by subtraction of the mean profile. Elevated intensities are observed for *Group II bulges* (e.g. C insertions at positions 11 to 15, 6 to 7 and 18 to 20 or G insertions at positions 4 to 5 and 7 to 8). A very distinct increase of the hybridization signal is observed for C insertions into the subsequence TCCCCT in the middle of the sequence. As shown in **(C)** *Group II bulges* (red markers) have significantly higher intensities compared to *Group I bulges* (blue markers).
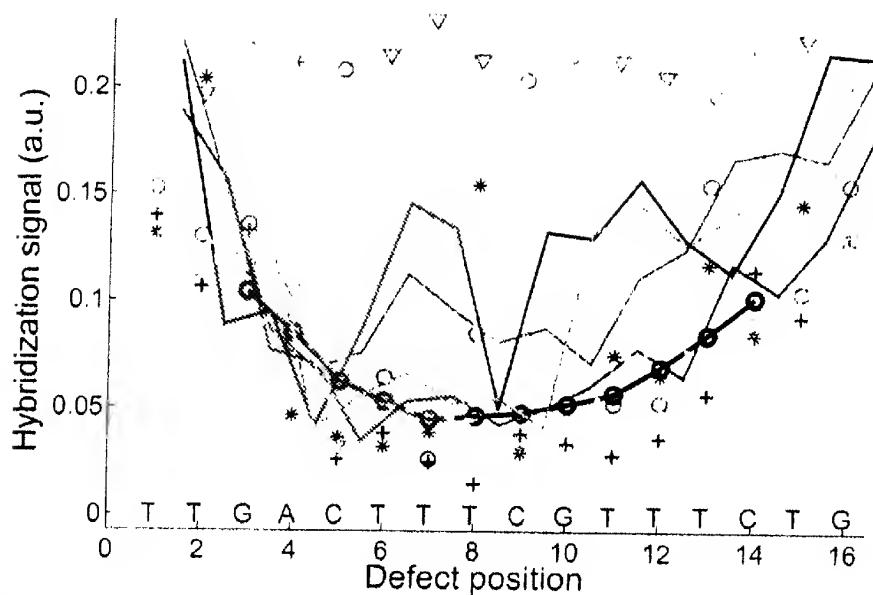
Figure 6
Direct comparison of single base mismatches, insertions and deletions. The 16 mer probe sequence motif 3'-TTGACTTTCGTTTCTG-5' is complementary to the target BEI. Hybridization signals (data processing: raw fluorescence intensities; solution-background correction) of single base mismatch probes with substituent bases A (red crosses), C (green circles), G (blue stars), T (cyan triangles), running average of mismatch intensities (black line); perfect match probe signals (grey symbols) single base insertion probes (solid lines) with insertion bases A (red), C (green), G (blue), T (cyan). Hybridization signals of single base deletions (orange dashed line) are comparable to that of mismatches at the same position. Increased hybridization signals of certain insertion defects are due to positional degeneracy of base bulges (see discussion).

nal data from 1000 different 20–25 mer probes indicates the general validity of the result

Interestingly, systematically increased hybridization signals (with respect to the averaged hybridization signal level from other defect types at the same position) have also been observed for certain Group I bulges: For G-insertions next to a T (e.g. in Fig. 5 at base position 15) we frequently find increased binding affinities similar to that of Group II bulges.

We further analyzed the degree of correlation between the binding affinities of probes with different insertion bases X and Y (see Additional file 5). A clear correlation appears between the hybridization signals of probes with T- and C-insertions, and also, though less distinct, between A-insertions. In contrast, we observed an anti-correlation between G- and A-insertions

DNA/DNA versus DNA/RNA mismatch and bulged hybridization
To investigate if the above results from DNA/DNA hybridization also apply to hybridization of RNA/DNA duplexes we performed a direct comparison employing DNA targets and corresponding RNA targets on the same microarray. We observed that MM discrimination in RNA/DNA duplexes is similar to MM discrimination in DNA/DNA duplexes (see Additional file 6). A statistical analysis (see Figs. 9 and 10) reveals, however, that purine-purine MMs are (with respect to the ranking order of MM stabilities, Fig. 10b,c and Fig. 10d) somewhat less stable in RNA/DNA duplexes than in DNA/DNA duplexes. The most sig-
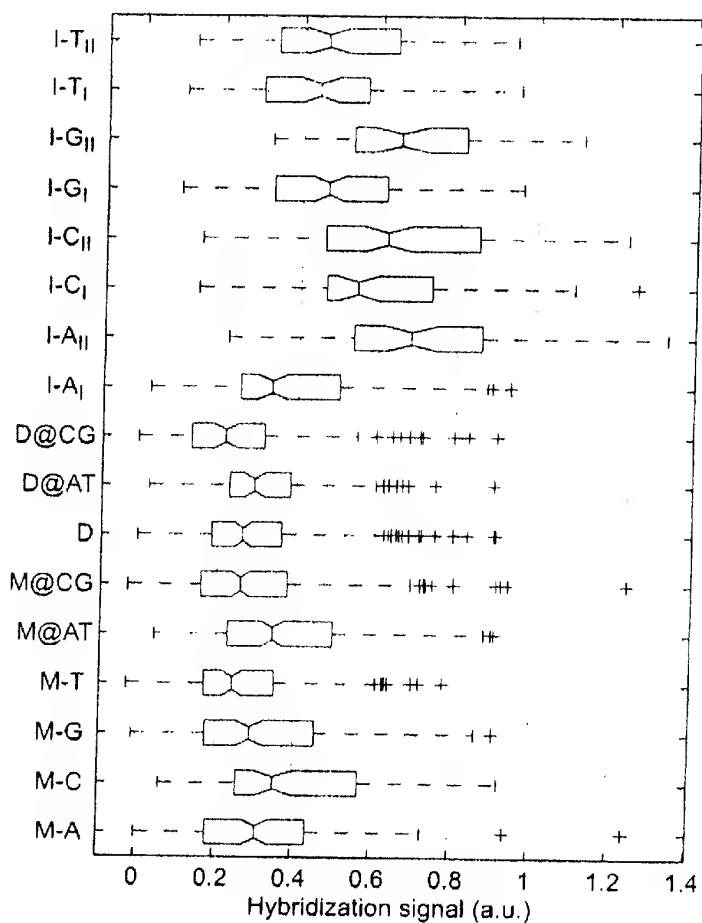
Figure 7
Comparison of the hybridization signals of different point mutation types. To minimize positional influence the statistics include only defect positions 5 to 12, located in the center of the 16 mer probes. The 1200 probe sequences were derived from 17 probe sequence motifs. Data processing: raw fluorescence intensity data; solution-background correction; hybridization signals are normalized by division by the corresponding perfect match hybridization signals. Defect categories: mismatch M-X (X: substituent base); mismatches at A·T and C·G sites M@AT, M@CG; single base deletion D; deletions at A·T and C·G sites D@AT, D@CG; single base insertion I-X$_{I/II}$ (X: insertion base, I/II: *Group I/Group II* base bulge). Hybridization signals from insertion probes (about 50% of the PM hybridization signal for *Group I*, 65% for *Group II* -median values) are significantly higher than that of MM probes (at about 30%). Mismatches at A·T sites result in about 25% larger hybridization signals than MMs at C·G sites. Deletion probes have a median hybridization signal that is slightly lower than the median MM hybridization signal. *Group I* base bulges with the exception of I-A$_I$ (33%) have hybridization signals of about 50% of the PM hybridization signal. Hybridization signals of *Group II* base bulges are (with the exception of T-insertions) significantly higher than that of the corresponding *Group I* bulges.
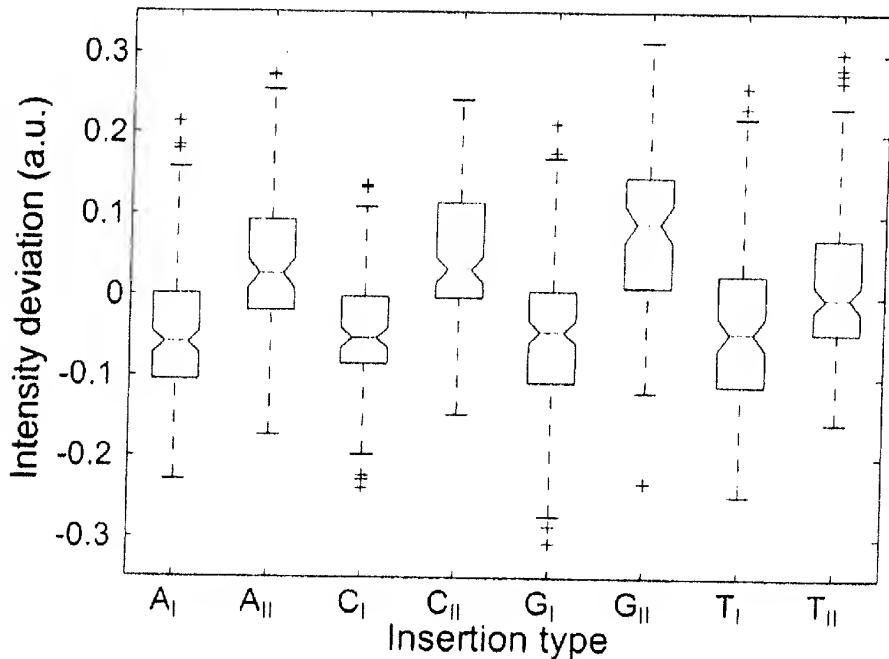
Figure 8
Boxplots show the hybridization signal deviations (from the mean profile) for the different insertion base types (A_I, C_I, G_I, T_I, A_{II}, C_{II}, G_{II}, T_{II}), which are differentiated according to affiliation to bulge *Group I/II*. Data processing: raw intensity data, solution-background correction; subtraction of the mean profile yields the defect-type dependent contribution of the hybridization signal. The statistical analysis includes about 1000 hybridization signals from 12 different 20 to 25 mer probe sequence motifs.

nificant differences between RNA/DNA and DNA/DNA MMs (see Additional file 7) are observed for the MM-types G·A and A·C (more stable in DNA/DNA duplexes) and for the MM-type T·G (which is more stable in RNA/DNA duplexes) A presumed destabilizing effect of purine-purine MMs in the ranking order of RNA/DNA mismatch discrimination (Fig. 10c) is superposed to the affected base pair effect (C·C vs. A·T – see above), which is very similarly, also observed in DNA/DNA hybridization.

For bulged duplexes we did not observe significant defect-type specific differences between RNA/DNA and DNA/DNA hybridization. The hybridization signal and sequence data from the microarray hybridization experiment are provided in Additional file 8.

Single base insertion, deletion and mismatch defects in comparison

Defect profiles for MMs and base bulges (Fig. 6) exhibit a very similar quantitative influence from defect position in DNA/DNA as well as in DNA/RNA complexes. For individual sequences the mean trough-shaped profile can be altered: Fig. 3 shows deformations of the trough-like profile on scales much larger than the size of a base pair.

Single base MM discrimination also depends on the type of MM base pair and the corresponding PM base pair (which has been substituted by the MM). Hybridization signals of MMs (normalized with the respect to the PM hybridization signal) originating from C·C base pairs are about 25% smaller (in the median) than for MMs from
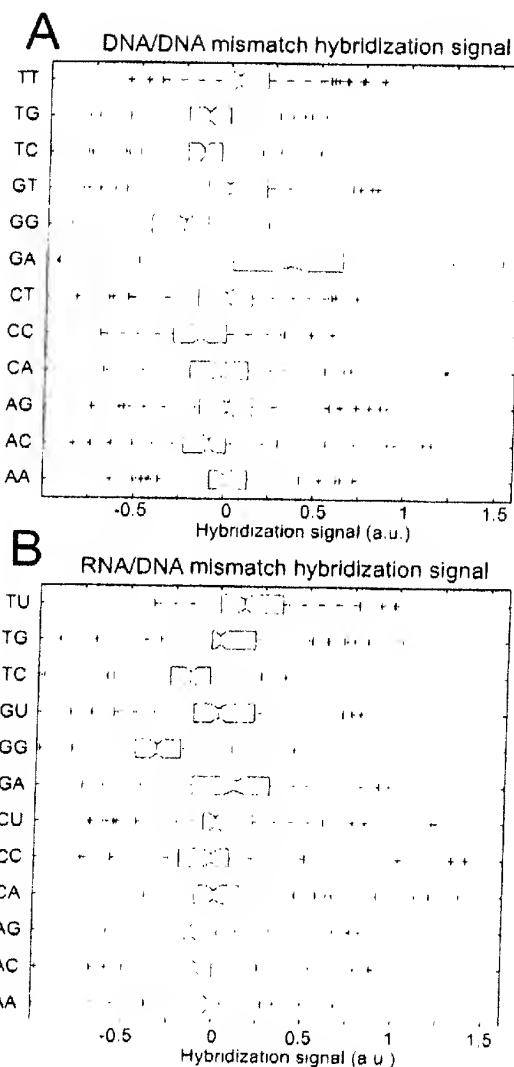
Figure 9
Comparison of DNA/DNA and RNA/DNA mismatch hybridization signals – statistical analysis. (A) MM-type related influence in DNA/DNA oligonucleotide duplexes. The positional influence was eliminated by subtraction of the moving average MM profile. Subsequent normalization was performed by division through the mean hybridization signal of the particular MM profile. (B) MM-type related influence in RNA/DNA oligonucleotide duplexes. Hybridization signal differences between the pairs of RNA/DNA- and analog DNA/DNA-duplexes are shown in Additional file 7.

a) DNA/DNA hybridization (large data set)

$$G \cdot A > T \cdot T \geq G \cdot T > A \cdot G \geq C \cdot T \approx C \cdot A > A \cdot A \approx T \cdot G \approx C \cdot C \approx T \cdot C \geq A \cdot C \geq G \cdot G$$

b) DNA/DNA hybridization (small data set for direct comparison with RNA/DNA hybridization)

$$G \cdot A > T \cdot T > A \cdot G \approx C \cdot T \approx G \cdot T \approx A \cdot A > C \cdot A > A \cdot C \geq T \cdot G \geq T \cdot C \approx C \cdot C > G \cdot G$$

c) RNA/DNA hybridization (small data set - equivalent to the DNA/DNA dataset in b)

$$T \cdot U \geq G \cdot A > T \cdot G \approx G \cdot U \approx C \cdot U \approx C \cdot A \approx C \cdot C \approx A \cdot A > A \cdot C \geq A \cdot G \geq T \cdot C > G \cdot G$$

d) Difference between RNA/DNA and DNA/RNA hybridization signals. Uracil is considered like thymine. (TG to GT: $I_{RNA/DNA} > I_{DNA/DNA}$; AC to GA: $I_{RNA/DNA} < I_{DNA/DNA}$)

$$T \cdot G > C \cdot A \geq C \cdot C \geq T \cdot T > C \cdot T \approx G \cdot T > A \cdot C \approx T \cdot C \geq A \cdot A \approx G \cdot G > A \cdot G > G \cdot A$$
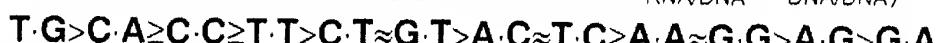
Figure 10
Comparison between DNA/DNA and RNA/DNA mismatch binding affinities. (a) Ranking order of DNA/DNA mismatch binding affinities (extracted from Fig. 4). (b) As anticipated the ranking order for DNA/DNA MMs obtained from the smaller subset of probe sequences (Fig. 9A) is very similar. The ranking order for the analogue RNA/DNA MM duplex stabilities (c) (extracted from Fig. 9B) reveals significant differences in comparison to (b). In part (d) MM-types are ordered according to the hybridization signal differences between RNA/DNA and DNA/DNA MMs (as extracted from Additional file 7) Purine-purine MMs (purine bases highlighted in blue) display the largest decrease of binding afinities with respect to other MM-types.

A T base pairs. Single base deletions affecting C·G base pairs result in about 30% smaller hybridization signals than deletions affecting A·T base pairs. The deletion profile in Fig 6 (orange dashed line) shows that the local ups and downs of the profile curve correlate with deletions affecting either A·T or C·G base pairs. Thus, for MMs and single base deletions it is the type of base pair affected by the point-mutation, which determines the impact on hybridization affinity to an important degree, however, it is still less important than defect-position

We also observe a noticeable influence of the next-neighbor bases of the mismatch (see Additional file 2)

Discussion
Dominating influence of defect position
We observe that defects located in the center of the oligonucleotide duplexes are significantly more destabilizing

than defects at the ends [10]. Similar influence of the MM position has been reported previously from other microarray based studies [7,9], and also – although sparsely – from hybridization experiments in solution [25,26] The limited data in solution may be due to the technical difficulty of studying a large number of different probes Quantitatively, in accordance with [7] we have identified MM position (relative to the duplex ends) as the strongest influential factor on the hybridization signal, when compared to MM-type and nearest neighbor effects

The well-established two-state nearest-neighbor model, which has proved to be reliable for the prediction of duplex stabilities in solution-phase, does not regard the position of the (mismatched) NN pairs [6] We propose that a model for the prediction of microarray binding affinities should also include the position of the NN pairs – in particular in case of mismatched NN pairs Affinity

models for microarray hybridization considering a positional dependence of the nearest-neighbor parameters have been previously discussed in [12,27-30].

We observe a very similar position dependence for single base bulge defects as for single base mismatches. Also, the magnitudes of the impacts of the MMs and base bulges on the hybridization signal are very similar (apart from the relative high binding affinity of Group II bulges). This consistency suggests a common origin of the positional influence, independent of defect type.

Sterical crowding at the surface, as suggested by Peterson et al. [31], can in principle reduce the accessibility of the probe surface-bound 3'-ends and can thus decrease the impact of defects located near this end. However, in our case we observe largely symmetrical intensity profiles with respect to both ends of the probes (Fig. 2).

Focusing on individual probe sequence motifs we observe, that the positional influence does not only depend on the defect-to-end distance, but also has a sequence-dependent contribution. This indicates that the impact of a defect also depends on the stability of the local sequence environment (beyond the nearest neighbors). Since there are no long range molecular forces, we infer that the molecular dynamics must play a role, effects like breathing bubbles or zipping could be at the origin. This influence of the duplex sequence and the observed symmetry of the defect positional influence with respect to the duplex ends suggest that end-domain opening (i.e. sequential unzipping of the double-helix from the duplex ends) must be suspected to be a key mechanism for understanding the influence of defect position on duplex stability.

Influence of the MM-type
Removing the positional influence in our data, we see that single-base MMs introduced at the site of a C·G base pair result in a larger decrease of the hybridization signal (with respect to the PM hybridization signal) than MM defects affecting A·T base pairs. The same applies for single base deletions (see Fig. 6). These experimental results (Fig. 4), in accordance with nearest-neighbor thermodynamic parameters for Watson-Crick base pairs [6], mainly reflect the increased base stacking and hydrogen bonding interactions of C·G base pairs. We observe a positive correlation between the experimentally determined single base mismatch discrimination and predicted free energy increments $\delta \Delta G_{37}^{\circ}$ (between MM and PM duplexes) on the basis of the nearest-neighbor model – for details see Additional file 2. A similar correlation (between $\log_2$(PM/MM)

hybridization signal values and $\delta \Delta G_{37}^{\circ}$) has been reported previously in [9].

We emphasize the good correlation between our DNA/DNA MM stability order (Fig. 12e) and the corresponding results of Wick et al. [9] (the MM stability order in Fig 12d was extracted from the plot of $\log_2$(PM/MM) hybridization signal values in Fig. 5a in [9]). A major difference, however, occurs for the MM-pair G·G, which is the least stable in our study. Wick (and also Sugimoto [32]) found G·G to be one of the most stable MMs. Interestingly, however, Pozhitkov et al [7] - in accordance with our results -identified G·G as one of the least stable MM-types.

Our direct comparison between DNA/DNA and RNA/DNA hybridization on microarrays reveals – for RNA/DNA duplexes – an increased destabilization of purine-purine mismatches, with respect to other MM types. An explanatory approach for the observed differences between DNA/DNA and RNA/DNA binding affinities is, that purine-purine MMs cause larger steric hindrance in the A-form hybrid duplexes than in the B-form DNA/DNA duplexes.

In contrast to [7] we did not observe that purine-purine mismatches in RNA/DNA duplexes are, in absolute terms, more discriminative than other MM-types

### Increased stability of Group II single base bulges
We observe significantly increased hybridization signals of single- base insertion defects in which the insertion base is placed next to a like-base. Our investigation shows that (on the microarray) the difference between Group I and Group II binding affinities δI_{bulge} (inferred from the hybridization signal I) is distinctly larger than the defect-type related variation of binding affinities δI_{MM} (see Fig 7). For comparison, the free energy differences among the MM trinucleotide duplexes $abc / \bar{a}x\bar{c}$ and $abc / \bar{a}y\bar{c}$ (mismatched bases x and y; neighboring bases a and b unchanged; overline denotes complementary bases) span the range $\delta \Delta G_{MM}^{37} = 0.5$ to 2.6 kcal/mol (calculated with MM nearest-neighbor free energies [33] for $T = 37 °C$).

The increased stability of Group II bulges in comparison with Group I bulges has been investigated previously in solution rather than on microarrays [24,34,35]. According to Ke and Wartell [34] the increased stability of Group II bulges originates from positional degeneracy of the base bulge. Additional conformational freedom, entailing higher entropy, results in lowered duplex free energy (thus in increased stability). According to Zhu et al [24] posi-

tion degeneracy accounts for an average stabilization of -0.3 to -0.4 kcal/mol (in agreement with the theoretical estimate [24] of $-R \cdot T \cdot \ln 2 = -0.43$ kcal/mol at 37°C) for a two-position degeneracy Znosko et al [35] reported Group II duplexes to be on average $\delta \Delta G^{37} = -0.8$ kcal/mol more stable than Group I duplexes. The latter value

matches better our observation of Group II hybridization close to the perfect match hybridization signal.

For explanation of the large binding affinity of Group II duplexes we propose the following mechanism (illustrated in Fig. 11) based on the molecular zipper model
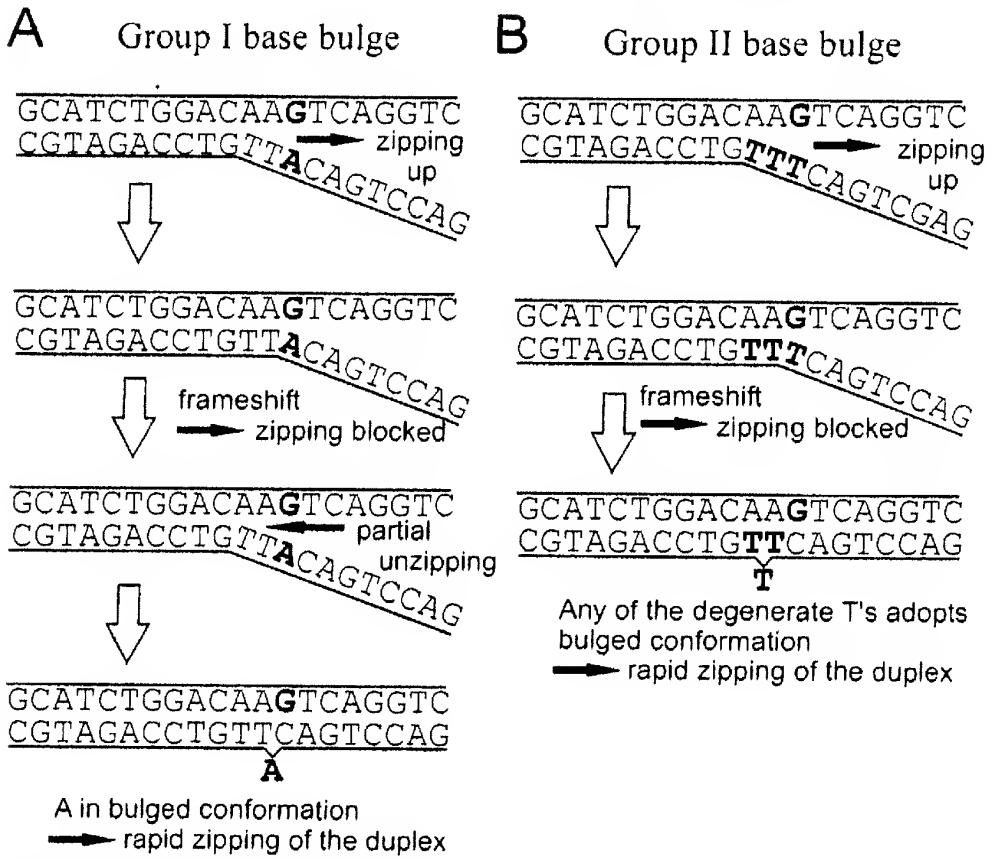
Figure 11
Proposed mechanism for the increased binding affinity of duplexes with Group II base bulges. The Group I base bulge (A), originating from the insertion of the unpaired base 'A', creates a 1-nt frameshift between the complementary probe and target sections, and thus acts like a barrier delaying the formation of a stable duplex. The bulged 'A' needs to adopt a favorable (e.g. looped out) conformation, so that the frameshift is compensated and the zipping of complementary base pairs can continue. Unlike the Group I base bulge in (A) the Group II base bulge in (B), originating from the insertion of the surplus base 'T' next to another 'T', is degenerate. Since there is an increased probability that any of the two degenerate bases adopts a favorable conformation, while simultaneously the subsequent base is forming a base pair with the corresponding base in the opposite strand (so that the frameshift is overcome), the formation of a stable duplex is accelerated.

[36,37] Even in thermal equilibrium due to thermal exci-lation. zipping (consecutive base pairing) as well as unzipping occur The extend of the end-domain denatur-ation of the duplex, which is described by a random walk (biased by the duplex sequence), may finally result in complete dissociation of the duplex. The binding affinity is determined by the ratio k_{nuc}/k_{diss} between the nucleation rate k_{nuc} and the duplex dissociation rate k_{diss}. We consider that a defect does not have an important influence on the unzipping, since the defect does not present a barrier for the process. For closing of the strands, however, the situa-tion is different. The surplus (bulged) base must act as a kinetic barrier, interrupting the rapid zipping of the duplex. The 1-nt frameshift between the (largely) comple-mentary strands, owing to the unpaired bulge base pre-vents closing beyond the defect and results in a partially zipped, and correspondingly weakly-bound, duplex. Duplex closure can only progress if the interfering surplus base is giving way (adopting a favorable looped-out or stacked conformation), thus allowing the subsequent base to form a Watson-Crick base pair with the comple-mentary base in the target strand. From this point zipping can progress rapidly Therefore compared to Watson-Crick nearest-neighbor pairs, a base bulge (similar to a mismatch) has a decreased ratio of zipping/unzipping-rates k_z/k_z and thus favors unzipping of the duplex (i e the duplex dissociation rate k_{diss} is increased with respect to the perfectly matching duplex). For *Group II* bulges the k_z/k_z ratio at the defect site is increased with respect to *Group I* bulges: in case of a *Group II* bulge there is an increased probability that any of the degenerate bases makes way (and adopts, for example, a favorable flipped-out confor-mation) while simultaneously the subsequent base forms a base pair This is due to the increased number of possi-ble molecular conformations, which can lead to continu-ation of the zipping. Then, as the frameshift is compensated, the rapid zipping to complete the duplex occurs. Since the nucleation rate k_{nuc} of *Group I* and *Group II* duplexes may be assumed to be the same, the binding affinity of *Group II* duplexes must be increased

a) Solution hybridization DNA/RNA (Sugimoto *et al.* 2000)

$$T \cdot G \gg G \cdot U \approx G \cdot G > G \cdot A \approx A \cdot G \approx C \cdot A > A \cdot A \approx T \cdot U \approx C \cdot U > A \cdot C \approx T \cdot C$$

b) Microarray hybridization DNA/RNA (Pozhitkov *et al.* 2006)

$$T \cdot G \approx T \cdot U \approx T \cdot C > G \cdot U \approx A \cdot C \approx C \cdot C \approx C \cdot U \approx A \cdot A \approx A \cdot G \geq C \cdot A > G \cdot G \approx G \cdot A$$

c) Gene silencing RNA/RNA (Schwarz *et al.* 2006)
Silencing efficiency depends on the single base mismatch between the mRNA and siRNA sequences

$$C \cdot A > U \cdot G \approx C \cdot U > U \cdot U > A \cdot C \approx G \cdot U \approx C \cdot C \geq U \cdot C > G \cdot A \geq G \cdot G \geq A \cdot A > A \cdot G$$

d) Microarray hybridization DNA/DNA (Wick *et al.* 2006)

$$G \cdot T \approx G \cdot A > T \cdot T > G \cdot G > T \cdot G \geq A \cdot G \approx C \cdot T \approx A \cdot A \geq C \cdot A \geq A \cdot C \geq C \cdot C \approx T \cdot C$$

e) Microarray hybridization DNA/DNA (this study)

$$G \cdot A > T \cdot T \geq G \cdot T > A \cdot G \geq C \cdot T \approx C \cdot A > A \cdot A \approx T \cdot G \approx C \cdot C \approx T \cdot C \geq A \cdot C \geq G \cdot G$$

Figure 12
Stability orders of MM-types X·Y for hybridization in solution (a) and on microarrays (b, d, e). In the microarray experiments (b, d and e) MM binding affinities have been normalized with the corresponding PM binding affinity, whereas the orders a) and c) reflect the absolute impact of the MM pairs on duplex binding affinity For the microarray MM-pairs (in b, d and e) the probe base X (DNA) is on the left and the target base Y (DNA or RNA) is on the right. The efficiency of RNA interfer-ence (c) (from [2]) is assumed to be determined by the stability of A-form RNA duplexes between the *RISC*-bound *guide strand* and the complementary mRNA. The left base X is part of the guide strand (at position 10) and the right base Y is part of the mRNA. Apart from the the base pair X Y the mRNA and siRNA sequences remained fixed. In (a) to (c) purine bases are high-lighted in blue. In (d) and (e) mismatches with respect to a perfect matching C G base pair are highlighted in red Details on the individual stability orders are provided in the text

Previous studies – Including RNA/DNA hybridization

Tautz and coworkers [7] performed a mismatch study with 20 mer oligonucleotide microarrays fabricated by light-directed *in situ* synthesis with the Geniom® One instrument (febit biomed GmbH, Heidelberg) Similar as in our study, they compared normalized hybridization signal intensities

However, an important difference between the experiments described in [7] and our experiments is the use of *in vitro* transcribed RNA targets [7] originating from ribosomal RNA They observe a more pronounced destabilization by purine-purine MMs compared to our results

A further study on the impact of MM stabilities in RNA/DNA duplexes, in solution rather than on a microarray surface, has been published by Sugimoto *et al.* [32] As discussed in [7] the destabilizing effect of purine-purine MMs is not observed by Sugimoto *et al.* [32]. However, the stability order in [32], referring to ΔG^{37} values of mismatched trinucleotide duplexes, is considering absolute stability parameters, whereas [7,9] and our study consider mismatch discrimination with the corresponding PM binding affinity as a reference level. Therefore, the comparability with the RNA/DNA stability order in [32] is limited A recent work on the impact of single base MMs in RNA-interference (RNAi) - allele-specific gene silencing experiments [2] - is interesting in the context of this study, since here the sequence recognition is based on base-pairing between the *guide strand* (a single RNA strand which is bound to the *RISC* complex) and a complementary mRNA. Schwarz *et al* (see Schwarz: *table 5b*) have shown that among all MM-types incorporated at position 10 of the guide strand (apart from the point mutations the sequence of the guide strand was preserved) purine-purine MMs resulted in the least silencing of gene activity (owing to a small binding affinity of the mismatched sequences), whereas U·G, C·U and U·U mismatches resulted in a very efficient gene silencing (see Fig. 12c). It is assumed that purine-purine MMs strongly interfere with the formation of an A-form helix between the guide strand and the target mRNA [38] This appears to be in accordance with the findings of Pozhitkov *et al.* on RNA/DNA MM discrimination. However, the inferred RNA/RNA mismatch stability order (shown in Fig. 12c) is not normalized with the corresponding PM stabilities, but rather reflects the absolute impact of the MM base pairs in a given duplex sequence and cannot be easily compared to our study and to [7].

Conclusion

We performed a comprehensive, array-based study on the influence of point defects on the binding affinity of oligonucleotide duplexes. Contrary to previous studies by others, we have employed well-defined hybridization

conditions by using short, end-labeled oligonucleotide target sequences (one at a time to minimize competitive effects) and can therefore exclude that target secondary structure, steric hindrance, labelling or competitive effects are relevant for an explanation of the observed results

In our microarray-based hybridization assays the binding affinity of mispaired duplexes is dominated by the influence of defect position The Influence of the defect-type is about half in magnitude, when compared to defect-position

There is also an Influence of the neighboring sequence, which has farther reach than the defect next neighbor Although this long reach interaction must somehow be related to the base stacking energies, we did not find a simple description. We attribute so far unexplained long range effects, in particular a trough-shaped position dependence, to molecular dynamics. We propose a molecular zipping mechanism as a suitable explanation Zipping agrees well with the observation that *Group II* bulges (bulges next to Identical bases) have stronger hybridization signals than expected from previous data. Experimentally, it is not completely clear, whether the strong positional influence on oligonucleotide binding affinity is restricted to surface-hybridization or if it is also relevant for solution-phase hybridization (maybe to a smaller extend) The comparison to other related work [2,7,32], however, shows significant differences in the MM-type dependence of duplex binding affinities. Our comparative analysis of the impact of point defects on the binding affinity of DNA/DNA and RNA/DNA duplexes reveals that purine-purine MMs are more destabilizing in the latter This may explain some discrepancies in the literature.

The use of DNA microarrays enables a detailed investigation of oligonucleotide duplex binding affinities producing a wealth of data in simple experiments. We demonstrate that important aspects (defect position influence, differences between DNA/DNA, RNA/DNA and RNA/RNA hybridization, surface and bulk hybridization) about the impact of point defects on oligonucleotide duplex binding affinities are not yet understood Our results from simple, controlled experiments agree well with results from extracting data from complex DNA target mixtures [7,9]. This shows that DNA hybridization on surfaces can be reproducible and quantitatively significant. Deviations from the behavior, which we describe here, are observed in microarray experiments and they must be due to complexity of DNA target mixtures.

applied for probes complementary to the target URA. At 30°C these, due to their large hybridization affinity, hybridize with reduced defect discrimination. Probes with a length of 20 and more bases are hybridized at 40°C. Hybridization is monitored in real-time on an Olympus IX81 fluorescence microscope. During acquisition of the hybridization signal the microarray is left in the hybridization solution. A 10 × 0.4NA UPlanApo objective provides a sufficiently large field of view. An electron multiplying CCD camera (Hamamatsu EM-CCD 9102) with a 1000 × 1000 pixel resolution is used for image acquisition During image acquisition shade correction is performed to compensate for intensity inhomogeneities in fluorescence excitation.

Image analysis software developed in-house is employed to read the intensities of hundreds of features simultaneously

Hybridization signal analysis – normalization

Hybridization signal measurements are performed with the microarray immersed in the hybridization solution Thus, the measured hybridization Intensity signal $I_{feat,meas}$ is composed of the feature intensity I_{feat} and the solution background intensity I_{back} (originating from fluorescent targets floating above the microarray in the hybridization solution) The overall intensity $I_{feat,meas} = f(x) \cdot (I_{feat} + I_{back})$ is affected by the function $f(x)$ which accounts for spatial variations of the fluorescence excitation and the light collection efficiency of the microscope system (e.g. due to vignetting). Apart from $I_{feat,meas}$ we also locally (i e. next to the corresponding microarray feature - see Additional file 10A) measure the solution background intensity $f(x) \cdot I_{back}$. A solution-background correction is performed by subtraction of the background fluorescence intensity. Further, by division by the solution background Intensity $f(x) \cdot I_{back}$ we cancel the feature-position related bias $f(x)$.

$$I_{feat,corr} = \frac{f(x) \cdot (I_{feat} - I_{back})}{f(x) \cdot I_{back}}. \qquad (1)$$

In the further analysis we separate between the relatively strong defect positional influence and the defect-type related influence on the binding affinity. The positional influence is calculated as the moving average of mismatch hybridization signals (including all mismatch types) over a window of five consecutive MM-positions By subtraction of the mean profile we obtain the MM-type dependent contributions δI_{MM} to the hybridization signal

To compare δI_{MM} from different defect profiles it is necessary to account for the fact that the mismatch discrimination depends on the binding affinity. Mismatch discrimination is stronger in weakly-binding short duplexes or duplexes with a large AT-content. Vice versa,

in case of duplexes with larger binding affinities the differences between PM and MM duplexes and among different MMs, respectively, may be rather small. We performed normalization of δI_{MM} by division by the standard deviation $\sigma_{profile}$ (see Additional file 10B), or, alternatively, by division by the average of all MM hybridization signals of the corresponding MM defect profile.

Design of the DNA chip experiments

The flexibility of the *in situ* synthesis and the excellent spot homogeneity simplifies a comprehensive comparative analysis with the capability to detect subtle differences of the probe binding affinities. The experiments mainly differ in selection and spatial arrangement of the probe sequences. Particular experiments focus on the extraction of the positional dependence, the comparison of different defect types and on the identification of further influential parameters

Spatial variations of the photodeprotection intensity and optical aberrations affecting the imaging contrast can result in gradients (as indicated in Additional file 11B) of the probe DNA quality (due to a varying number of synthesis errors). Thus, for a reliable determination of subtle differences in hybridization affinities, probes to be compared directly should be closely spaced on the microarray

In the following we describe the design of the individual experiments:

Single base mismatch study

To investigate the positional dependence of single base mismatches and the impact of the mismatch type, we designed microarrays containing comprehensive sets of MM probes derived from a series of 25 16 mer probe sequence motifs. Position and type of the mismatch base pair were systematically varied, allowing us later to distinguish between the dominating positional dependence and other influential factors.

The features are arranged in groups of four (see Additional file 11A), corresponding to the four possible substituent bases (A, C, G and T) at a particular base position A group comprises three mismatch probes plus one perfect match probe (PM) used for control Sixteen of these feature groups (one for each base position) are arranged in a square feature block comprising in total 64 features (Additional files 9 and 11A)

Single base bulges

Single base insertions and deletions, due to an extra unpaired base result in bulged duplexes with reduced stability A comprehensive study on the Impact of single base insertions was performed using the chip design shown in Additional file 11A The experiment comprised about

1000 single base insertion probes (insertion base type and position systematically varied) derived from twelve 20 to 25 mer probe sequence motifs.

Direct comparison of single base MMs and single base bulges
An experiment allowing for a direct comparison of PM, MM, single base insertion and deletion probes has been performed. Probe sets were derived from 16 mer probe sequence motifs, complementary to the targets listed in Tab 1. For each of the 16 possible defect positions a set of 9 probes (comprising four single base insertions, one base deletion, three MMs and one PM probe) has been created. To avoid that a regular arrangement of the probe features could possibly affect the measurement (e.g. by introducing a bias due to increased target depletion near a PM probe), the sets of nine probes were randomly arranged in 3 × 3 matrices (Additional file 11B).

Direct comparison between DNA/DNA and DNA/RNA mismatches
The chip design (Additional file 11B) and the experimental procedures were basically identical with that of the previous experiment. Hybridization assays were conducted with fluorescently labeled DNA targets and corresponding RNA targets (Tab 1). To avoid fabrication-related variation of the hybridization signals the hybridization assays were performed on the same chip, initially with RNA and subsequently, after regeneration of the microarray (by heating to 70°C in pure hybridization buffer), with the corresponding DNA targets

Three microarrays were fabricated, each one focussing on one particular target sequence (COM, PET and LBE). Each microarray assay investigated single base MM and bulge defects for 6 different probe sequence motifs (obtained by shifting the 16 to 20 mer probe motif with respect to the longer target sequence). Two replicates of each feature block are employed to control for the reproducibility of the measurement

Hybridization assays with the three microarrays were performed independently and on different days. The subsets of data obtained from the each of the assays display the same defect-type dependent trend for the defect-type dependent binding affinities. Yet smaller subsets from the individual defect profiles (originating from a single probe sequence motif) show basically the same trend of binding affinities which is, however, superposed by a strong sequence dependent bias

Authors' contributions
TN developed the experimental setup, performed the experiments, carried out the data analysis and drafted the manuscript OE aided in DNA chip synthesis and data analysis TM participated in the development of the DNA microarray synthesizer WM and JK participated in data

interpretation and helped to draft the manuscript. AO conceived of the study, and participated in its design and coordination and aided in drafting the manuscript

Additional material

Additional file 1
The raw hybridization signal intensities of the 16 mer probes in a microarray hybridization experiment on single base mismatch discrimination The data was extracted from fluorescence micrographs (16-bit gray scale TIFF images) of the hybridized microarrays. The dataset comprises the hybridization signal raw data and probe/target sequences of 24 mismatch defect profiles.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S1.txt]

Additional file 2
Correlation between the MM-type related hybridization signal deviations from the mean profile δI_{mp} and the predicted Gibbs free energy increments $\delta \Delta G_{37}^{\circ}$ between MM and corresponding PM duplexes $\delta \Delta G_{37}^{\circ}$ was calculated from mismatch NN-parameters [4] Hybridization signal data processing as described in Additional file 10 The MM-type is categorized according to the MM base pair X·Y (in A) and according to the flanking base pairs (in B). Data points indicate the median δI_{mp} of the individual mismatch/flanking base pair categories The small number of data within the individual categories (owing to the combinatorial increase of mismatch/nearest neighbor categories) can result in outliers The exact MM-category corresponding to each data point can be identified by looking up the symbols in the identical plots in (A) and (B) Part (A) shows a weak, approximately linear correlation between δI_{mp} and $\delta \Delta G_{37}^{\circ}$, indicating that the MM discrimination on microarrays can be related to MM nearest-neighbor parameters established from solution-phase hybridization A relatively weak mismatch discrimination can be observed for a variety of MM-types with $\delta \Delta G_{37}^{\circ} < 1.5$ kcal/mol Part (B) indicates that mismatch-types with C·G-flanking base pairs at both sides have (on average) larger hybridization signals than mismatches with A·T flanking base pairs at both sides. Among the more stable MM-types above the trend line (which serves the purpose to split each of the individual MM base pair type related clusters – shown in part A – in two halves) 20 have C·G-only flanking pairs and 19 have A·T-only flanking pairs In contrast, among the less stable MM-types – below the trend line – only 13 have C·G-only flanking pairs, whereas 29 have A·T-only flanking pairs
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S2.eps]

Additional file 3
The raw hybridization signal intensities of the 22-26 mer probes in a microarray hybridization experiment on the binding affinity of bulged duplexes Hybridization signal intensities were extracted from fluorescence micrographs (16-bit gray scale TIFF images) of the hybridized microarrays The dataset comprises the hybridization signal data and probe/target sequences of 14 defect profiles
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S3.txt]

Additional file 4

The raw hybridization signal intensities of the 16 mer probes in a microarray hybridization experiment designed for a direct comparison between binding affinities of single base mismatches and single base bulges. The data was extracted from fluorescence micrographs (16-bit gray scale TIFF images) of the hybridized microarrays. The dataset comprises the hybridization signal raw data and probe/target sequences of 23 defect profiles.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S4.txt]

Additional file 5

Histograms of hybridization signal differences IX-IY (X and Y denote the different insertion bases in otherwise identical probe sequences) reveal correlations between the hybridization signals of different insertion types. To exclude the impact of systematically increased intensities of Group II insertions only Group I insertions are regarded here. Between T- and C-insertions (and between C- and A- insertions) a correlation, as indicated by a narrow distribution with a pronounced peak near zero, is observed. The broad distribution of hybridization signals differences between C and A insertions doesn't show a distinct peak, indicating that there is no correlation but rather an anti-correlation for insertions of A and C.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S5.eps]

Additional file 6

Comparison of DNA/DNA and RNA/DNA mismatch hybridization signals – mismatch defect profiles. Parts A–D make a direct comparison of hybridization signals, obtained from subsequent hybridization of RNA targets (top) and DNA targets (bottom) on the same microarray. The defect positional influence is identical for DNA/DNA and RNA/DNA hybridization. However, the impact of MM-types reveals systematic differences. The sequences shown in the plots are the probe sequence motifs that have been modified by base substitution. The hybridization signal (in a.u.) is plotted against the defect position. Hybridization signal processing: solution background correction (see Methods section). Substitution bases A (red cross), C (green circle), G (blue star) and T (cyan triangle) either result in 3 MM duplexes and one PM duplex at every defect position. Hybridization signals of duplexes with single base deletions (yellow line), moving average MM hybridization signal (black line).
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S6.eps]

Additional file 7

Hybridization signal variation between pairs of mismatched RNA/DNA- and analog DNA/DNA- duplexes (hybridization signals of DNA/DNA duplexes were subtracted from the hybridization signals of the corresponding RNA/DNA duplexes). The largest differences between RNA/DNA and DNA/DNA binding affinities were found for the MM-types T-C, G-A and A-G.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S7.eps]

Additional file 8

The raw hybridization signal intensities of this microarray hybridization experiment provide a direct comparison between RNA/DNA and DNA/DNA hybridization. Hybridization signal intensity raw data was extracted from fluorescence micrographs (16-bit gray scale TIFF images) of the hybridized microarrays. The dataset contains the data of 3 independent experiments (performed with 3 different microarrays). Each microarray dataset comprises the hybridization signal data and probe/target sequences of 24 defect profiles.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S8.txt]

Additional file 9

Fluorescence micrograph of hybridized features (feature size 21 µm) in the 16 mer mismatch experiment. The shading-corrected image shows two feature blocks corresponding to two different 16 mer probe sequence motifs (3'-TTCAGCCATATTACTG-5' – to the left, 3'-TATTACTGCACCT-CAC-5' – to the right) both hybridizing with the fluorescently labeled target sequence COM (5'-Cy3-AACTCGCTATAATCACCTGCACTG-3'). Each feature block comprises all single base mismatches of the particular probe sequence. Groups of four features (as indicated by the marked groups 1 and 2) correspond to each one of the 16 possible mismatch base positions. As indicated by the letters between the feature blocks the uppermost row of features in each group corresponds to an A base at the corresponding base position, followed by probes with C, G and T (see also Additional file 11A). The brightest feature within each group corresponds to the perfect matching probe. Nonhybridized targets in the hybridization solution contribute to the background intensity between the features. Mismatch intensity profiles for the probe sequence motif 3'-TATTACTGCAC-CTCAC-5' are shown in Fig. 2.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S9.eps]

Additional file 10

Data analysis procedures. (A) To reduce intensity gradients on the microarray (bias described by the spatially varying function f(x)) originating from the fluorescence microscope optics (e.g. due to inhomogeneous fluorescence excitation or vignetting) we apply a bias correction procedure on the raw intensity data: The $I(x)$ component in the raw hybridization signal intensity is canceled by normalization with the local solution background fluorescence intensity $I(x)$ I_{back}. (B) Normalization of the MM-type dependent component δI_{MM} of the hybridization signal is necessary since the magnitude of mismatch discrimination depends on the binding affinity of the sequence motif. The defect profile in a) shows a large MM discrimination (typical for a weakly bound duplex), whereas the defect profile in b) shows a small MM discrimination (typical for more strongly bound duplexes). In the position-independent defect profiles (right) the positional influence (obtained as the moving average of all MM types over five consecutive defect positions – shown as a bold line in the defect profile in the left image) has been subtracted, to yield the MM-type dependent influence δI_{MM}. For statistical analysis of the defect-type contribution, including comparable data from different defect profiles, normalization is performed by division through the standard deviation $\sigma_{profile}$ of the position-independent defect profile.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S10.eps]

Additional file 11

Microarray feature arrangements (A) for the single base mismatch/single base insertion experiments (compare with Additional file 9) For the direct comparison between single base MMs and single base bulges and for the comparison of DNA/DNA and RNA/DNA hybridization the feature arrangement (B) was used This more compact arrangement of features has been chosen to minimize the impact of gradient effects on the relative hybridization signal values of the various defect types. The 9 features belonging to each defect position (depicted with dashed boxes for positions 1 and 16) comprise 3 single base MMs, 4 single base insertions, one single base deletion and one perfect matching probe. The gradient indicated in (B) demonstrates that the erroneous variation within the closely spaced feature set belonging to a particular defect position is significantly smaller than for features located far apart
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6750-8-48-S11 eps]

References

1. Conner BJ, Reyes AA, Morin C, Itakura K, Teplitz RL, Wallace RB: Detection of sickle-cell beta-s-globin allele by hybridization with synthetic oligonucleotides. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 1983, 80:278-282.
2. Schwarz DS, Ding HL, Kennington L, Moore JT, Schelter J, Burchard J, Linsley PS, Aronin N, Xu ZS, Zamore PD: Designing siRNA that distinguish between genes that differ by a single nucleotide. *Plos Genetics* 2006, 2(9):e140.
3. Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K: Hybridization of synthetic oligodeoxyribonucleotides to phi-chi-174 DNA effect of single base pair mismatch. *Nucleic Acids Research* 1979, 6(11):3543-3557.
4. Allawi HT, SantaLucia J: Thermodynamics and NMR of internal GT mismatches in DNA. *Biochemistry* 1997, 36(34):10581-10594.
5. Peyret N, Seneviratne PA, Allawi HT, SantaLucia J: Nearest-neighbor thermodynamics and NMR of DNA sequences with internal AA, CC, GG, and TT mismatches. *Biochemistry* 1999, 38(12):3468-3477.
6. SantaLucia J, Hicks D: The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure* 2004, 33:415-440.
7. Pozhitkov A, Noble PA, Domazet-Loso T, Nolte AW, Sonnenberg R, Staehler P, Beier M, Tautz D: Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Research* 2006, 34(9):e66.
8. Urakawa H, El Fantroussi S, Smidt H, Smoot JC, Tribou EH, Kelly JJ, Noble PA, Stahl DA: Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays. *Applied and environmental microbiology* 2003, 69(5):2848-2856.
9. Wick LM, Rouillard JM, Whittam TS, Gulari E, Tiedje JM, Hashsham SA: On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes. *Nucleic Acids Research* 2006, 34(3):e26.
10. Naiser T, Mai T, Michel W, Ott A: Versatile maskless microscope projection photolithography system and its application in light-directed fabrication of DNA microarrays. *Review of Scientific Instruments* 2006, 77(6):063711
11. Luebke KJ, Balog RP, Garner HR: Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to RNA transcripts. *Nucleic Acids Research* 2003, 31(2):750-758.
12. Binder H: Thermodynamics of competitive surface adsorption on DNA microarrays. *Journal of Physics-Condensed Matter* 2006, 18(18):S491-S523
13. Bishop J, Blair S, Chagovetz AM: A competitive kinetic model of nucleic acid surface hybridization in the presence of point mutants. *Biophysical Journal* 2006, 90(3):831-840.
14. Fodor SPA, Read JL, Pirrung MC, Land Lu Stryer AT, Solas D: Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991, 251(4995):767-773.
15. McGall GH, Barone AD, Diggelmann M, Fodor SA, Gentalen E, Ngo N: The efficiency of light-directed synthesis of DNA arrays on glass substrates. *Journal of the American Chemical Society* 1997, 119(22):5081-5090.
16. Singh-Gasson S, Green RD, Yue YJ, Nelson C, Blattner F, Sussman MR, Cerrina F: Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nature Biotechnology* 1999, 17(10):974-978.
17. Gao XL, LeProust E, Zheng H, Srivannavit O, Gulari E, Yu PL, Nishiguchi C, Xiang Q, Zhou XC: A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Research* 2001, 29(22):4744-4750.
18. Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg JP, Ballin J, McCormick M, Norton J, Pollock T, Sumwalt T, Butcher L, Porter D, Molla M, Hall C, Blattner F, Sussman MR, Wallace RL, Cerrina F, Green RD: Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Research* 2002, 12(11):1749-1755.
19. Luebke KJ, Balog RP, Mittelman D, Gerner HR: Digital optical chemistry: A novel system for the rapid fabrication of custom oligonucleotide arrays. In *Microfabricated Sensors, Application of Optical Technology for DNA Analysis Volume 815* Edited by: Richard Kordal, Author Usmani, Wei Tak Law American Chemical Society Publications; 2002:87-106.
20. Cerrina F, Blattner F, Huang W, Hue Y, Green R, Singh-Gasson S, Sussman M: Biological lithography: development of a maskless microarray synthesizer for DNA chips. *Microelectronic Engineering* 2002, 61-2:33-40.
21. Baum M, Bielau S, Rittner N, Schmid K, Eggelbusch K, Dahms M, Schlauersbach A, Tahedi H, Beier M, Guimil R, Scheffler M, Hermann C, Funk JM, Wikmerten A, Rebscher H, Honig M, Andreae C, Buchner D, Moschel E, Glathe A, Jager E, Thom M, Greil A, Bestvater F, Obermeier F, Burgmaier J, Thome K, Weichert S, Hein S, Binnewies T, Foitzik V, Muller M, Stahler CF, Stahler PF: Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. *Nucleic Acids Research* 2003, 31(23):e151.
22. Suzuki S, Ono N, Furusawa C, Kashiwagi A, Yomo T: Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays. *BMC Genomics* 2007, 8:373.
23. Fish DJ, Horne MT, Brewood GP, Goodarzi JP, Alismayehu S, Bhendiwad A, Searles RP, Benight AS: DNA multiplex hybridization on microarrays and thermodynamic stability in solution: x direct comparison. *Nucleic Acids Research* 2007, 35(21):7197-7208.
24. Zhu J, Wartell RM: The effect of base sequence on the stability of RNA and DNA single base bulges. *Biochemistry* 1999, 38(48):15986-15993.
25. Kierzek R, Burkard ME, Turner DH: Thermodynamics of single mismatches in RNA duplexes. *Biochemistry* 1999, 38(43):14214-14223.
26. Dorris DR, Nguyen A, Gieser L, Lockner R, Lublinsky A, Patterson M, Touma E, Sendera TJ, Elghanian R, Mazumder A: Oligodeoxyribonucleotide probe accessibility on a three-dimensional DNA microarray surface and the effect of hybridization time on the accuracy of expression ratios. *BMC Biotechnology* 2003, 3:6
27. Naef F, Magnasco MO: Solving the riddle of the bright mismatches: Labeling and effective binding in oligonucleotide arrays. *Physical Review E* 2003, 68:011906.
28. Zhang L, Miles MF, Aldape KD: A model of molecular interactions on short oligonucleotide microarrays. *Nature Biotechnology* 2003, 21(7):818-821.
29. Mei R, Hubbell E, Bekiranov S, Mittmann M, Christians FC, Shen MM, Lu G, Feng J, Liu WM, Ryder T, Kaplan P, Kulp D, Webster TA: Probe selection for high-density oligonucleotide arrays. *Proceedings of the National Academy of Sciences of the United States of America* 2003, 100(20):11237-11242.
30. Binder H, Kirsten T, Loeffler M, Stadler P: Sequence specific sensitivity of oligonucleotide probes. *Proceedings of the German Bioinformatics Conference 2003* 2003, 2:145-147.

31. Peterson AW, Wolf LK, Georgiadis RM: **Hybridization of mis-matched or partially matched DNA at surfaces.** *Journal of the American Chemical Society* 2002, 124(49):14601-14607.

32. Sugimoto N, Nakano M, Nakano S: **Thermodynamics-structure relationship of single mismatches in RNA/DNA duplexes.** *Biochemistry* 2000, 39(37):11270-11281.

33. Allawi HT, SantaLucia J: **Nearest neighbor thermodynamic parameters for internal GA mismatches in DNA.** *Biochemistry* 1998, 37(8):2170-2179.

34. Ke SH, Wartell RM: **Influence of neighboring base-pairs on the stability of single-base bulges and base-pairs in a DNA frag-ment.** *Biochemistry* 1995, 34(14):4593-4600.

35. Znosko BM, Silvestri SB, Volkman H, Boswell B, Serra MJ: **Thermo-dynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with single nucle-otide bulges.** *Biochemistry* 2002, 41(33):10406-10417.

36. Gibbs JH, Dimarzio EA: **Statistical mechanics of helix-coil tran-sitions in biological macromolecules.** *Journal of Chemical Physics* 1959, 30:271-282.

37. Kittel C: **Phase transition of a molecular zipper.** *American Jour-nal of Physics* 1969, 37(9):917-&.

38. Rodriguez-Lebron E, Paulson HL: **Allele-specific RNA interfer-ence for neurological disease.** *Gene Therapy* 2006, 13(6):576-581.

39. Le Berre V, Trevisiol E, Dagkessamanskaia A, Sokol S, Caminade AM, Majoral JP, Meunier B, Francois J: **Dendrimeric coating of glass slides for sensitive DNA microarrays analysis.** *Nucleic Acids Research* 2003, 31(16):e88.

40. Kim C, Li M, Venkataramaia N, Richmond K, Kaysen J, Cerrina F: **DNA microarrays: an imaging study.** *Journal of Vacuum Science & Technology B* 2003, 21:2946.

41. Hasan A, Stengele KP, Giegrich H, Cornwell P, Isham KR, Sachleben RA, Pfleiderer W, Foote RS: **Photolabile protecting groups for nucleosides: Synthesis and photodeprotection rates.** *Tetrahe-dron* 1997, 53(12):4247-4264.

Exhibit 3:   Naiser et al., "Position Dependent Mismatch Discrimination of DNA Microarrays
– Experiments and Model," BMC Bioinformatics 9:509 (2008)

BMC Bioinformatics

Research article

Position dependent mismatch discrimination on DNA microarrays – experiments and model

Thomas Naiser*[1], Jona Kayser[1], Timo Mai[1], Wolfgang Michel[1] and Albrecht Ott[1,2]

Address [1]Experimentalphysik I, Universität Bayreuth, D-95440 Bayreuth, Germany and [2]Experimentalphysik, Universität des Saarlandes, D-66041 Saarbrucken, Germany

E-mail: Thomas Naiser* - thomas.naiser@googlemail.com; Jona Kayser - jona.kayser@ep1.uni-bayreuth.de; Timo Mai - timo.mai@ep1.uni-bayreuth.de; Wolfgang Michel - wolfgang.michel@ep1.uni-bayreuth.de; Albrecht Ott - albrecht.ott@physik.uni-saarland.de
* Corresponding author

Abstract

Background: The propensity of oligonucleotide strands to form stable duplexes with complementary sequences is fundamental to a variety of biological and biotechnological processes as various as microRNA signalling, microarray hybridization and PCR. Yet our understanding of oligonucleotide hybridization, in particular in presence of surfaces, is rather limited. Here we use oligonucleotide microarrays made in-house by optically controlled DNA synthesis to produce probe sets comprising all possible single base mismatches and base bulges for each of 20 sequence motifs under study.

Results: We observe that mismatch discrimination is mostly determined by the defect position (relative to the duplex ends) as well as by the sequence context. We investigate the thermodynamics of the oligonucleotide duplexes on the basis of double-ended molecular zipper. Theoretical predictions of defect positional influence as well as long range sequence influence agree well with the experimental results.

Conclusion: Molecular zipping at thermodynamic equilibrium explains the binding affinity of mismatched DNA duplexes on microarrays well. The position dependent nearest neighbor model (PDNN) can be inferred from it. Quantitative understanding of microarray experiments from first principles is in reach.

Background

The well-known double-helix structure of nucleic acids results from sequence-specific binding between complementary single strands. Sequential base pairing between A-T and C-G base pairs along the two complementary strands results in the formation of stable duplexes. This so called hybridization process is fundamental to many biological processes and biotechnologies. Microarrays consist of surface-tethered *probe* sequences, which act as specific scavengers for their respective complementary *target* sequence. The molecular recognition enables a highly parallel detection of nucleic acid sequences in complex target mixtures. Hybridization also occurs with single mismatched (MM) base pairs, however, these duplexes are significantly less stable than the corresponding perfect match (PM) [1, 2]. The single base pair mismatch-discrimination capability of short (~20 nt) oligonucleotide probes provides an important diagnostic

tool for the detection of point-mutations and single nucleotide polymorphisms (SNP's) [3]. DNA duplex stability arises from hydrogen bonding and base stacking interactions (the latter comprise van der Waals interactions, electrostatic and hydrophobic interactions between adjacent base pairs) According to the well-established nearest-neighbor model, thermodynamically a nucleic acid duplex can be considered the sum of these nearest-neighbor (NN) interactions [4-6]. The binding free energy of an oligonucleotide duplex can be predicted from the nearest-neighbor free energy parameters. The helix propagation parameters (one for each of the 10 possible base-pair doublets in case of a DNA/DNA duplex) account for the duplex sequence. Further parameters provide corrections for duplex initiation, A·T terminal pairs or a symmetry penalty in case of self-complementary sequences. The NN model adequately predicts oligonucleotide duplex melting temperatures T_M in bulk solution [7]. Datasets of Watson-Crick NN parameters [8] provide the basis for nucleic acid structure and melting temperature prediction software like the DINAMelt web server [9] (UNAFold), the HYTHER server and others. The NN model can be extended beyond the Watson-Crick pairs to include single base MM defects [7, 10].

In spite of good knowledge about nucleic acid hybridization in solution, the prediction of binding affinities on DNA microarrays remains empirical. Recent microarray studies [11-15] report, that the influence of even a point defect on hybridization signal intensity cannot be predicted easily. In particular the influence of defect position on the hybridization signal is stronger than the influence of MM-type [12, 14, 16].

Experiments show that the two-state nearest-neighbor (TSNN) approach [7], which has been very successful in predicting duplex stability in solution, does not appropriately describe MM binding affinities on DNA microarrays. The NN model does not account for the position of the individual NN pairs [7], except for the outermost ones. Based on microarray data, Zhang et al [17] proposed a position dependent nearest-neighbor (PDNN) model. The model assumes that the duplex binding free energy can be expressed as a weighted sum of stacking energies with empirically derived positional weight parameters [17-21]. The purpose of this study is to investigate the influence of point defects on (surface bound) hybridization experimentally and theoretically. Previous studies investigate mismatch discrimination with samples of very different sequence motifs [11, 12]. However, other effects such as secondary structure formation or competitive binding may reduce the visibility of the impact of the MM-defect on the binding affinity. To avoid such complications we performed

experiments with fixed sequence motifs. We focus on small variations of the probe sequences. We perform hybridization studies with home-made microarrays comprising sets of very similar probe sequences. We use a single target sequence in each hybridization assay in order to avoid inter-target binding as well as target competition of different sequences for one and the same probe sequence. In order to avoid excluded volume interactions or secondary structure we limit the length of the target sequence to be of the order of the probes. These simplifications (described in detail in [14]) enable a detailed investigation of the influences of defect type, defect position, flanking base pairs and the sequence motif on the binding affinity. The extensive set of hybridization affinities obtained from our experiments enables us to perform a very complete analysis. We compare the experimental data to theoretical modeling based on a double-ended molecular zipper approach (the double-ended nucleic acid zipper has been previously described by [22-26]). We find that in order to reproduce the microarray hybridization signal in our model, the heterogeneity of binding affinities - mostly owing to in situ synthesis-related probe defects (e.g. probe polydispersity) - needs to be taken into account. More than that, synthesis defects arise as useful for parallel detection of many different sequences.

Methods

DNA Microarray Hybridization Experiments

Hybridization assays are performed on high-density oligonucleotide microarrays (see Fig. 1). These microarrays (DNA Chips) are fabricated in-house [14] on the basis of light-directed solid-phase combinatorial chemistry [27, 28]. A "maskless" photolithographic technique [13, 29-31] based on a digital micromirror device type spatial light modulator (DMD™, Texas Instruments Inc.) enables tailor-made design of DNA microarrays (with up to 25000 different probe sequences) on a laboratory scale. Point defects - single base substitutions, insertions and deletions - are produced in the in situ synthesis process by variation of the nucleotide coupling scheme for the particular probe sequence.

Protocols for the preparation of dendrimer-functionalized microarray substrates (adapted from [32]) and for the light-directed synthesis (based on NPPOC-phosphoramidites [33]), as well as details on the hybridization assay and on fluorescence microscopy based microarray analysis (Fig. 1) are provided in Naiser et al [14, 15].

In each microarray hybridization assay a probe set of cognate probes with purposefully introduced point mutations - derived from a common probe sequence
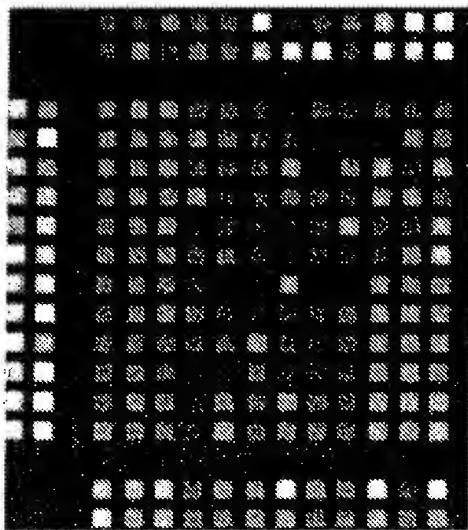
Figure 1
Fluorescence micrograph (taken with an Olympus IX81 epi-fluorescence microscope and a Hamamatsu EM-CCD camera) of a microarray feature-block comprising variations of the 16 mer probe sequence motif 3'-TATTACTGGACCTGAC-5'. Microarray hybridization was performed with the 5'-Cy3-labeled RNA oligonucleotide target 3'-AACUCGCUAUAAUGACCUGGACUG-5' (target concentration: 1 nM In 5 × SSPE, pH 7.4, 0.01% Tween-20, $T = 30°C$). Each $3 × 3$ sub-array comprises (randomly arranged) one perfect matching probe, three single base mismatch probes, four insertion probes and one single base deletion probe. In Fig. 2A the hybridization signals (fluorescence intensities, averaged over the center of the microarray features) are plotted versus the defect position. The size of each microarray feature is 21 μm and the pitch of the array is 35 μm. The significantly brighter feature-block at left comprises variations of the 20 mer probe sequence motif 3'-TTGAGCGATATTACTGGACC-5'.

motif – is hybridized against a single target sequence, which perfectly matches the probe sequence motif. We systematically vary defect type and defect position to provide the complete "defect profile" of hybridization affinities with probe sets We include not only all single base mismatches (MMs), but also, in order to investigate mismatch discrimination in a broader context of other sequence defects, we consider single base bulges (originating from insertions and deletions) as well as probes with multiple defects. Since the ca 130 probes within each probe set differ only by single bases we are able to distinguish between defect-positional and sequence influence In our experimental conditions hybridization equilibrium is reached after a few tens of minutes Further details can be found in [14].

Results

Position Dependent Influence of Single Base MMs and Bulges on Probe-Target Binding Affinity

From the fluorescence micrograph (Fig. 1), we obtain the hybridization signals, which we plot as a function of defect position (Fig. 2). We note a strong influence of the defect position on probe-target binding affinity which is larger than the influence of the defect type We find that bulge defects display a very similar position-dependent influence on hybridization signal intensity to mismatches Furthermore we observe that the magnitude of mismatch discrimination (and bulge discrimination) at a particular defect position (i e the shape of the defect profile) depends on the duplex sequence.

As can directly be inferred from Fig. 2, defects in the middle of the probes are most destabilizing. In the center of a 16 mer duplex a single nucleotide MM typically reduces the hybridization signal to 0–40% of the corresponding PM duplex hybridization signal. Defect type and nearest-neighbor effects have less influence on the hybridization signal than defect position. Our experiments show a mostly monotonous decrease of hybridization signals over a range of typically 5–8 defect positions (for 16 mer probes and up to 14 positions for some 25 mer sequence motifs) from the duplex ends towards the center of the duplex. This is consistent with previous work [11, 12]

In order to separate the defect positional influence (DPI) for a particular probe sequence motif from the defect type related influences we run a moving average filter on the defect profile. We observe that the DPI is not only a simple function of the distance between the defect and the duplex-ends, but it is also related to the nucleotide sequence (compare Fig. 2A and 2B and Fig. 3A and 3B)

We also perform hybridization experiments on oligonucleotide duplexes with two single base deletion defects at varying positions x and y The results show that the binding affinity depends also on the relative position of the defects (for details see Additional file 1, Fig. S5 and [26]) The hybridization signal is largest if each defect is located close to an end Lowest binding affinities are observed for defect configurations which divide the sequence into three roughly equally long subsequences Closely spaced defects (with a distance of less than four nucleotides) systematically increase their impact with distance

A

B
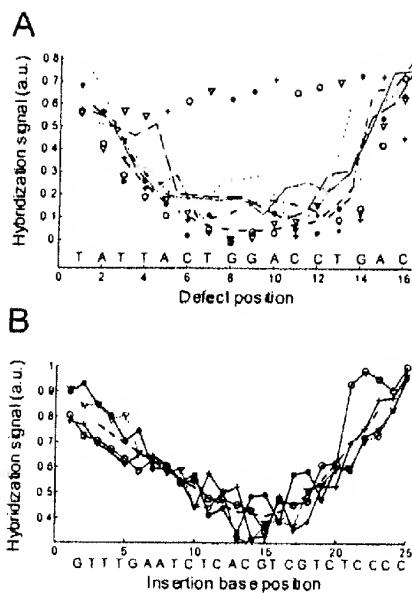
Figure 2

The "defect profile" shows the position-dependent impact of single base mismatches, insertions and deletions on hybridization affinity. Symbols: MM probes with substituent bases A (*red crosses*), C (*green circles*), G (*blue stars*), T (*light blue triangles*); moving average of all MM intensities (*black dashed curve*); single base insertion probes with insertion bases A (*red dash-dotted curve*), C (*green solid curve*), G (*blue dotted curve*), T (*light blue dashed curve*). Defect profiles of different probe sequence motifs. (A) Position dependent impact of various single base defects on the hybridization affinity for the probe sequence motif 3'-TATTACTGGACCTGAC-5' (hybridized with the complementary RNA target sequence). Hybridization signals of single base deletions (*orange dashed curve*) are comparable to that of MMs at the same position. PM probe signal replicates (*black symbols on gray ground*) serve as an indicator for spatial bias on the microarray. Deviations of MM hybridization signals from the mean profile are mostly MM-type specific. Increased hybridization signals of certain insertion probes (where the bulged surplus base is located next to identical bases – Group II bulges [14, 53]) are due to positional degeneracy of the bulge defects. (B) Position dependent impact of various single base insertion defects on the hybridization affinity for the probe sequence motif 3'-GTTTGAATCTCACGTCGTCTCCCC-5' (hybridized with the complementary DNA target sequence). Insertions of A (*red crosses*), C (*green circles*), G (*blue stars*), T (*light blue triangles*), moving average of all insertion probe intensities (*black dashed curve*). Systematically increased hybridization signals of Group II bulges are discussed in Additional file 1, Fig. S2.

Discussion

Single base mismatches and base bulges alike show a strong, trough-shaped position-dependent influence biased by the considered sequence motif. Experimental evidence for an influence of the sequence context (beyond the nearest neighbors) on the stability of single base pair MMs has been reported previously (hybridization of short 31 bp linear oligonucleotide duplexes in bulk solution) by Benight and coworkers [34], however such effects have not yet been systematically quantified. The commonly used two-state model of nucleic acid hybridization between the microarray probe P and the target strand T resulting in the formation of the duplex D is described by Eq 1

$$P + T \quad \overset{k_{nuc}}{\underset{k_{diss}}{\rightleftharpoons}} \quad D \qquad (1)$$

In thermodynamic equilibrium duplex nucleation (determined by the slow nucleation rate k_{nuc}) is balanced by duplex dissociation with the dissociation rate k_{diss}. The widely used two-state nearest-neighbor model (including mismatched NN-dimers as described by [10]) cannot provide an explanation for this positional influence, it does not account for the position of the individual nearest-neighbor dimers. We assume that the nucleation rates k_{nuc} of very similar duplexes (differing by a single base pair, e.g. a PM duplex and a corresponding mismatched duplex) are virtually identical. Thus, the positional dependence observed experimentally can be expected to result from differences in k_{diss}. In agreement with [25] we show that the positional influence originates from end-domain unzipping. Our experimental findings suggest a common mechanism for DPI, that is independent of the defect type. Further, the relatively long range of the DPI (Fig 3A and 3B) suggests that molecular dynamics may well be a good candidate for an explanation. The symmetry of DPI (with respect to the duplex ends) and sequence-specific deviations from the symmetry indicate a zipping related mechanism. Thus, in order to account for partial denatured duplex states, we use a double-ended zipper model of the oligonucleotide duplex to determine mismatched oligonucleotide duplex stabilities as a function of defect position. We consider a situation in thermodynamic equilibrium.

Double-ended Zipper Model

We check if a double-ended zipper model [22-25] (Fig. 4), considering end-domain-denaturation only, is appropriate to describe the experimental observations. Internal denaturation, due to the large bubble initiation barrier (owing to stacking interactions towards both sides of a nucleotide) and due to the relatively short length of the duplexes, is expected to be negligible [22].
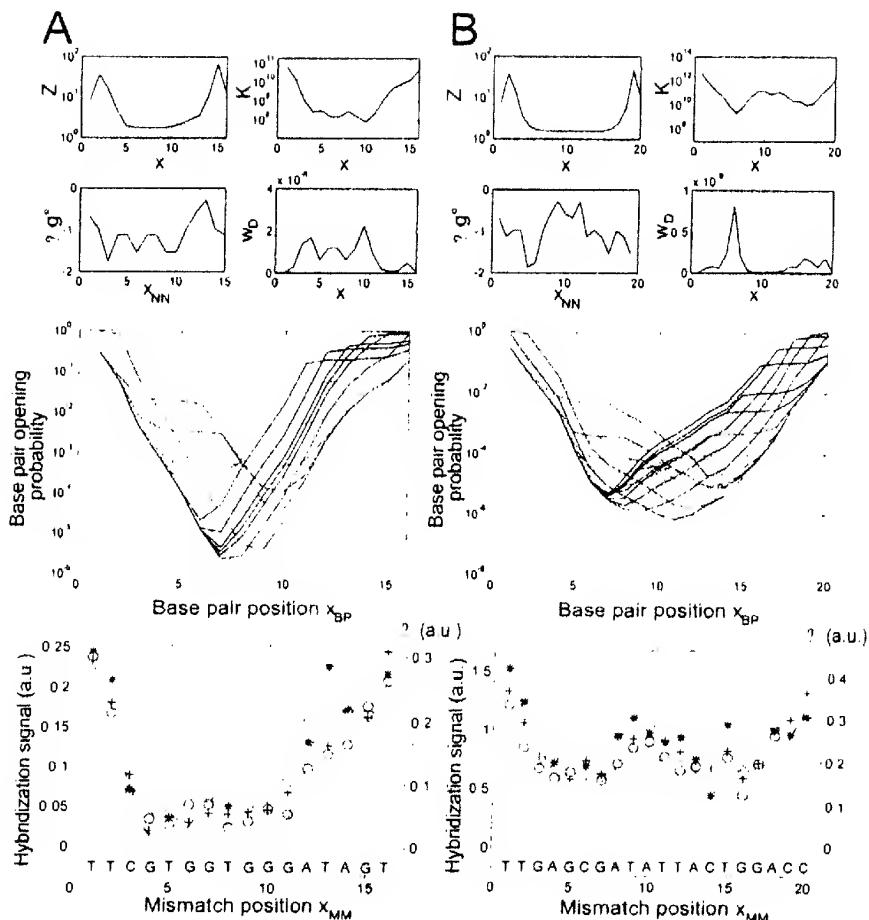
Figure 3

Comparison of simulation results with the experimentally determined hybridization affinities for two probe sequence motifs (A) and (B). The four small sub-figures in the top section (from *top left* to *bottom right*) show the partition function Z and the duplex binding constant K as a function of defect position x (semi-logarithmic plots), the NN-free energies Δg° of particular NN-pairs as a function of NN-pair position x_{NN}, and the statistical weight for complete duplex dissociation w_D as a function of defect position. Irregularities in Z(x) at the duplex ends are an artifact caused by the fact that only a single NN-pair is affected by a MM-base pair at the duplex end. The middle sub-figure shows the base pair opening probabilities (the fraction of strands in which the corresponding base pair at position x_{BP} is unzipped) as a function of the defect position. The spectrum of differently colored curves encodes the different defect positions x_{MM} (*red* – defect at left end; *purple* – defect at right duplex end). The bottom sub-figure compares the experimentally determined MM defect profile (mismatched base: A (*red cross*), C (*green circle*), G (*blue star*), T (*cyan triangle*), *gray symbols* correspond to PM probes) with the simulated MM defect profile $\theta(x)$ (*dashed orange line*). With $\Delta g_{def} = 1$ kcal/mol (at the simulation temperature of 325 K) and an error rate of 12 percent (per synthesis step) the calculated defect profile $\theta(x)$ matches well the experimental data.
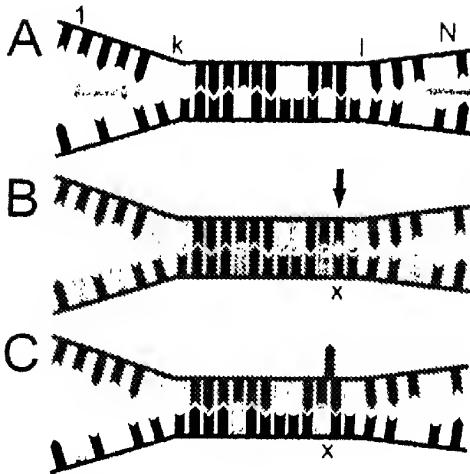
Figure 4
**Double-ended zipper model of the oligonucleotide
duplex.** (A) Unzipping of the relatively short duplexes is
initiated at the duplex ends only. The end-domain opening,
which progresses back and forth (nucleotide by nucleotide)
in a stepwise, zipper-like fashion, can be considered a biased
random walk. The energy level of the partially denatured
hybridization state $S_{k,l}$ (with respect to the completely
hybridized ground state) is determined by summation over
the NN free energies of the unzipped NN-pairs (from 1 to k
and from l to N). (B) Single base MMs (non-Watson-Crick
base pairing affect the stabilities of two adjacent NN-pairs at
positions x and $x + 1$. (C) Base insertions and deletions result
in bulged duplexes with an unpaired base. The surplus base
(depicted in a looped out conformation), similar as a MM
defect, results in a reduced binding affinity.

Using a partition function approach the impact of point
defects is investigated at thermodynamic equilibrium. We
perform this analytically, independently of a particular
sequence, as well as numerically with sequence-depen-
dence – using unified NN-parameters [8].

According to Craig et al [35] a kinetic scheme describing
helix growth and dissociation is given in Eq. 2.

$$P + T \quad \overset{k_+}{\underset{k_-}{\rightleftharpoons}} \quad D_1 \quad \overset{k_+}{\underset{k_-}{\rightleftharpoons}} \quad D_2 \quad \overset{k_+}{\underset{k_-}{\rightleftharpoons}} \quad \cdots \quad \overset{k_+}{\underset{k_-}{\rightleftharpoons}} \quad D_{n-1} \quad \overset{k_+}{\underset{k_-}{\rightleftharpoons}} \quad D_n \quad (2)$$

k_+ and k_- are the fast zipping and unzipping rates
determined by the nearest-neighbor propagation
parameters of the individual base-pair doublets. The
time-evolution of the oligonucleotide zipper can be
considered a biased random walk with a finite

probability for complete dissociation (described by the
duplex dissociation rate k_{diss}). Since we consider thermo-
dynamic equilibrium, we can use a partition function for
fast numerics.

Partition Function Approach (PFA) to Investigate
Oligonucleotide Duplex Thermodynamics
We use a partition function approach [22-25] and
investigate if the double-ended zipper model can
reproduce our experimental results. On the basis of
unified NN-Parameters [8] we calculate statistical
weights of partially denatured duplex states. The effect
of partial binding with respect to microarray data was
discussed earlier in [24-26].

The partition function Z_D of the duplex (Eq. 3) is the
sum of the statistical weights $w_{k,l}$ of all partially
hybridized duplex states $S_{k,l}$ (see Fig. 4).

$$Z_D = \sum_{k=0}^{N-1} \sum_{l=k+1}^{N} w_{k,l} = \sum_{k=0}^{N-1} \sum_{l=k+1}^{N} e^{\Delta G^*_{k,l}/RT} \quad (3)$$

The statistical weight $w_{k,l}$ of the partially denatured state
$S_{k,l}$ is calculated from the sum $\Delta G^*_{k,l}$ of NN free energies
Δg^*_i of the unzipped duplex sections (Eq. 4). $\Delta G^*_{k,l}$ can
be considered as the free energy level of the partially
denatured state.

$$\Delta G^*_{k,l} = \sum_{i=1}^{k} \Delta g^*_i + \sum_{i=l+1}^{N} \Delta g^*_i$$

$$\Delta G^*_{0,l} = \sum_{i=l+1}^{N} \Delta g^*_i \quad \Delta G^*_{k,N} = \sum_{i=1}^{k} \Delta g^*_i \quad (4)$$

NN free energies of Watson-Crick NN-pairs are deduced
from unified NN parameters [7]

$$\Delta g^*_i = \Delta h^*_i - T \cdot \Delta s^*_i \quad (5)$$

For the completely dissociated duplexes we estimate
partitions functions of probes Z_P and targets Z_T as

$$Z_P = Z_T = e^{\Delta G^*_D/(2RT)} \quad \Delta G^*_D = \sum_{i=1}^{N} \Delta g^*_i \quad (6)$$

For simplicity duplex initiation free energies have been
neglected here. Based on the duplex sequence we can
now calculate the duplex binding constant

$$K = \frac{Z_D}{Z_P Z_T} = \frac{Z_D}{e^{\Delta G^0_D/RT}} \quad (7)$$

Consideration of Point Defects

We introduce a defect parameter Δg^*_{def} (a simplified description of the mismatch NN parameters in [7, 10]) to account for the point defect at the defect position x (a similar approach is described in [25]).

An analytical derivation of the DPI for homopolymer sequences shows that the partition function (provided as a function of defect position - see Eq. 8) is increased for defects located near the duplex ends.

$$Z_D(x) = Z_{D_{PM}} + \left(e^{\frac{(N-x)\Delta g^\circ}{RT}} + e^{\frac{x\Delta g^\circ}{RT}} \right)\left(e^{\frac{\delta\Delta g^\circ_{def}}{RT}} - 1 \right)$$

(8)

In Eq 8 the defect impact $\delta\Delta g^*_{def} = \Delta g^*_{def} - \Delta g^\circ$ has been factored out, revealing a general (defect-type independent) position dependence that is largely governed by the distance between the defect at position x from the duplex ends. Defects proximate to the duplex ends increase end-domain opening. The partition function is increased due to the number of thermally populated (partially denatured) duplex states. The defect destabilization $\delta\Delta g^*_{def}$ determines how far Z_D is elevated in respect to the perfect match partition function ($Z_{PM} \approx 1$) and thus how far the DPI propagates into the interior of the duplex. With Eqs 7 and 8 we obtain an expression for the DPI on the duplex binding constant $K(x)$

$$K = \frac{\left(e^{\frac{x\Delta g^\circ}{RT}} + e^{\frac{(N-x)\Delta g^\circ}{RT}} \right)\left(e^{\frac{\delta\Delta g^\circ_{def}}{RT}} - 1 \right) + 1}{\left(e^{\frac{N\Delta g^\circ}{RT}} \right)\left(e^{\frac{\delta\Delta g^\circ_{def}}{RT}} \right)}$$

(9)

Fig. 5 illustrates Eq 9 for two different duplex stabilities.

While defects near the duplex ends result in low mismatch discrimination only (i e small reduction of K with respect to the PM binding affinity) defects in the center result in higher MM discrimination as K then approaches the value of the two-state equilibrium constant. NN-pair free energy increments $\delta\Delta g^*_{def_i}$ for single base MMs are in the range of 1 to 3 kcal/mol per NN pair (derived from NN parameters [8, 10]).
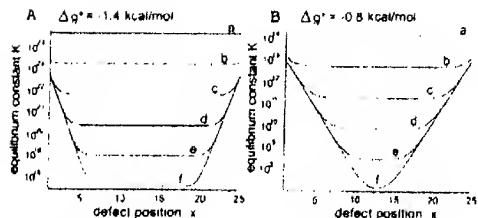
Figure 5
Positional influence of single base MM defects on the duplex stability for two different NN pair free energies Δg° **at a temperature of 310 K.** Curves a to f correspond to defect destabilization values $\delta\Delta g^*_{def}$ of 0 to 5 kcal/mol (incrementally increased by 1 kcal/mol). Defect destabilization $\delta\Delta g^*_{def}$ is quoted per affected NN pair. (A) $\Delta g^\circ = -1.4$ kcal/mol, this corresponds to an average NN-pair free energy; (B) $\Delta g^\circ = -0.8$ kcal/mol corresponding to a weakly bound sequence of A·T and T·A base pairs.

Employing these values in Eq 9 for $\Delta g^\circ = -1.4$ kcal/mol (Fig. 5A), DPI propagation is restricted to 3 or 6 NN-pairs, respectively. However, in subsequences with weakly bound NN-pairs (as demonstrated in Fig. 5B) the DPI can propagate further towards the middle of the duplex.

Relation Between the Hybridization Signal and the Binding Free Energy ΔG_D

In order to compare our numerical analysis to the experimentally observed hybridization signals we need to understand how the hybridization signal (fluorescence intensity from hybridized targets) is linked to duplex stability. As detailed below the assumption of a single (homogeneous) binding affinity within a microarray feature of the Langmuir adsorption model does not describe the experimentally observed hybridization signal intensities well. In this section we account for the heterogeneity that is introduced by in situ synthesis related random mutations of the microarray probe sequences

The importance of the adsorption model for the description of microarray hybridization has been discussed previously in [36-39] In the simplest description the equilibrium between single stranded probes and targets and hybridized duplexes $T + P \rightleftharpoons D$ can be described by a Langmuir-type adsorption isotherm (Eq 10) Under our experimental conditions targets were in sufficient excess, the target concentration $[T] \approx [T_0]$ can be taken as constant Since the hybridization signal intensity is expected to be proportional to the fraction of

hybridized probes $\theta = |D|/|P_0|$ we will in the following refer to θ as the hybridization signal.

$$\theta = \frac{|D|}{|P_0|} = \frac{K \cdot |T_0|}{1 + K \cdot |T_0|} \qquad (10)$$

Taking $K = e^{-\Delta G_h / RT}$ we obtain a sigmoidal relation between the hybridization signal and duplex free energy ΔG_D

$$\theta = \frac{e^{-\Delta G_D^0 / RT} \cdot |T_0|}{1 + e^{-\Delta G_D^0 / RT} \cdot |T_0|} \qquad (11)$$

Our experimental data suggest an approximately linear relation between the hybridization signal and the duplex binding free energies (within the free energy range covered by the defect profiles). However, with Eq. 11 an approximately linear relation between θ and ΔG_D is only provided within a narrow range $\delta \Delta G^* \approx 6$ kcal/mol (at $T = 310$ K and $|T_0| = 1$ nM). This cannot reproduce the experimentally observed DPI of the hybridization signal, since the free energy range of the defect profile exceeds the transition region. To investigate how the fluorescence intensity of hybridized targets is related to duplex stability on the microarray surface we performed a hybridization assay comprising sets of probes in which the probe length (assumed to be roughly proportional to duplex free energy) is incrementally increased (inset in Fig. 6). The experimental results in Fig. 6 show a sigmoid relation between the hybridization signal and probe length. However the transition region extends over at least 13 base pairs ($\delta \Delta G_{D37}^* \approx 20$ kcal/mol) over which a monotonous increase of the hybridization signal is observed. In agreement with our findings a linear relation between microarray hybridization signals (on spotted microarrays) and duplex binding free energies ΔG_D^* (derived from calorimetric measurements) has been reported recently by Fish et al. [40]. The large deviation from the Langmuir-equation agrees with previous observations [25, 41]. An effective isotherm with a broadened transition region, a Sips-isotherm, has been reported [25, 42, 43] to provide a better description of surface hybridization on microarrays. This isotherm can result from a heterogeneous, gaussian distribution of binding affinities. Reasons given for the heterogeneity include variation of the probe local environment, surface electrostatics [44] and entropic blockage [45]. As we show in the following a major contribution to the heterogeneity of binding affinities is probe polydispersity [25, 38, 46, 47], which is a result of sequence defects generated in the in situ synthesis process of DNA Chips, which introduces single base mismatches, base bulges and truncations
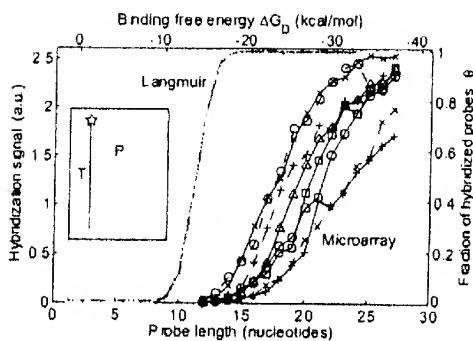
Figure 6
Hybridization signal versus duplex stability. The sigmoid transfer function $\theta(\Delta G_D^*)$ of the Langmuir isotherm (*right scale*) has a narrow transition region ($\delta \Delta G_D^* \approx 6$ kcal/ mol at a temperature of 310 K and a target concentration of 1 nM). Microarray hybridization signals (*left scale*) for incrementally increased duplex stabilities: The probe sequence motif was translated along the target sequence in increments of two bases (see inset), thus providing a set of different curves. All probes were hybridized with the common target sequence URA (1 nM in 5 × SSPE, for 20 minutes at 45°C). The approximately linear increase of the hybridization signal in the transition region extends over at least 13 base pairs ($\delta \Delta G_{D37}^* \approx 20$ kcal/mol).

Assuming a stepwise error rate of 10%, more than 90% of the 25 mer duplexes contain at least one synthesis error [13]. Since the number of synthesis errors per probe follows a binomial distribution, the majority of the strands contains between one and three single base defects.

We calculate binding constants K_i of the individual, randomly "mutated" probe sequences on the basis of the zipper model. Using the approach of Forman et al. [48] we obtain the total hybridization signal by summing up over the distribution of probes, where the contribution of each individual mutated probe θ_i is described by a Langmuir equation (Eq. 10) with the binding constant K_i. Probe polydispersity (in length as well as in sequence) reproduces a "stretched isotherm" [47] (similar to a Sips isotherm), with a significantly broadened transition region. This explains our experimental results in Figs. 6 and 2 well. A simulation of the transfer function $\theta(\Delta G_D^*)$ for various error rates and a comparison between the experimental data in Fig 6 and the corresponding simulation results are provided in Additional file 1, Fig. S6

Numerical Analysis of Mismatched Duplex Stability–Comparison with Experimental Results

To model experimental results with the partition function approach we choose the NN free energy of the mismatched base pair $\Delta g'_{def}$ as a free parameter $\Delta g'_{def} = 1$ kcal/mol (at $T = 325$ K) describes our experimental observations (in particular the dominating positional influence with respect to defect type-related influences) best – see Fig. 3 This value is also in good agreement with bulk solution parameters [10]. Results of the numerical simulation (in Fig. 3A) demonstrate that the shallower slope of the hybridization signal at the right duplex-end corresponds to a series of weak NN pairs (as anticipated by Eq 9) The partition function $Z(x)$ largely determines the positional influence. Additionally, as shown in Fig. 3B, defect-type related influence (the difference between MM and PM free energies $\Delta\Delta g°$ affects the statistical weight of the completely dissociated duplex w_D) is reflected in the hybridization affinity $K(x)$ and in the hybridization signal $\theta(x)$ In addition to single base pair defects our binding affinity model reproduces well our experimental results on the binding affinities of oligonucleotide duplexes with two single base deletion defects (for details see Additional file 1, Fig. S5)

In Fig. 7 we investigate the influence of heterogeneous probe-target binding affinities (see previous paragraph) on the shape of the defect profile. If the range of the mismatched duplex free energies is within the transition
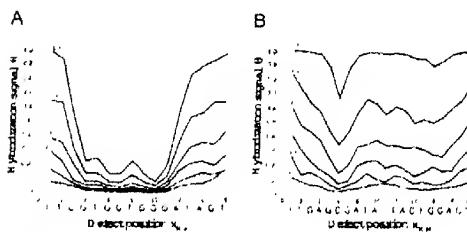
Figure 7
Influence of the synthesis error rate on the shape of the single base mismatch defect profile. The defect profiles (which correspond to the experimental data in Fig. 3 (A) and (B)) were calculated for error rates between 0 and 20 percent (per nucleotide coupling step). In (A) a positional influence is rather independent of the error rate – the duplex free energy range covered by the defect profile is within the approx. linear transition region. Whereas in (B) at an error rate of 0 percent, the free energy range of the defect profile doesn't match the transition region – the positional influence is hardly visible. At larger error rates the positional influence becomes dominating over the defect-type related influence.

region we observe an approximately linear relation between the hybridization intensity and the binding free energy |40| If the defect profiles free energy range exceeds the narrow transition region (like for example Fig. 7B, at an error rate of 0 percent) the positional influence remains hardly visible.

Approximation of the Zipper Model with a Position Dependent Nearest-Neighbor (PDNN) Model

In order to investigate the generality of our finding, we investigate if PDNN models, which fit experimental data well, can be inferred from our model framework. We note that zippering has been previously proposed as the rationale behind the PDNN model in |25|

In the following we investigate the contribution of each base pair to duplex stability and ask if there is a position-dependent contribution of Watson-Crick NN pairs in the same way as for defects.

This idea is the basis of the PDNN model [17, 21, 41] in which $\Delta G'_D$ is obtained as a position-dependent weighted sum of nearest-neighbor free energies.

$$\Delta G'_D = \sum_{i=1}^{N} w_i \Delta g'_i \qquad (12)$$

Following our theoretical approach we create a set of 7500 oligonucleotide duplexes assembled from a given set of NN pairs Although the TSNN (two-state nearest-neighbor) free energy of these duplexes is identical, the calculation with the zipper model indicates significant differences among the stabilities of the individual duplex sequences (see Additional file 1, Fig. S7) We investigate the positional distribution of NN pairs in the weakest/strongest 5% of the duplexes We find that in the most stable duplexes the stronger NN-pairs are located in the center whereas in the least stable duplexes the strong NN-pairs are located near the duplex ends. This result has been reproduced with the partition function based UNAfold software (DINAMelt web server [9]) with excellent agreement to the zipper model. A similar investigation (see Fig. 8) employing a set of random duplexes composed of nonidentical NN-pairs confirms the result. In Additional file 1, Fig S8 we show that duplex free energy values determined with the zipper model can indeed be approximated with a PDNN model The positional weights – described by a parabolic function $w_i(x)$ – have their maximum in the middle of the duplex. The results in Figs. 8 and Additional file 1, Fig S8 indicate that the contribution of the outer NN-positions to duplex stability decreases with increasing temperature. At 340 K the three outermost NN pairs (which is in total six of 24 NN pairs) have a significantly
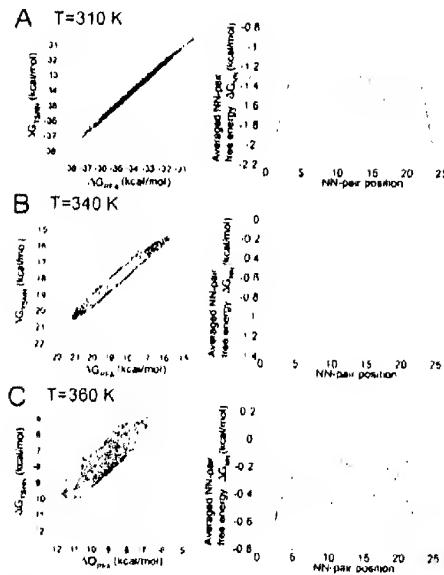
Figure 8
Comparison of the two-state nearest neighbor (TSNN) model and the zipper model (partition function approach – PFA). To investigate for which sequences the difference between TSNN free energies and PFA free energies is largest, we have created a large set of 5000 random 25 mer sequences with a similar nucleobase composition. Scatter plots of TSNN free energies versus PFA free energies (left) show a very good correlation at a temperature of 310 K (A). At higher temperatures (340 K and 360 K shown in B and C) we find significant deviations between the two models. We have selected the 5% of sequences with the largest residuals (highlighted by red symbols) and determined the position-dependent distribution of NN free energies (shown right) by averaging (→ averaged NN pair free energy versus NN-pair position). The Gibbs free energies in upper, middle and lower plots refer to temperatures T = 310 K, 340 K and 360 K, respectively). At 310 K the sequences with the most stable ΔG_{PFA} have their weak NN pairs at the outermost two base positions (dashed black line) and therefore the more strongly binding NN pairs in the interior. Vice versa sequences with the weakest ΔG_{PFA} (solid green line) have strong NN pairs located at the outermost positions. The mean NN free energy (average over all sequences) is indicated by the dotted red line. At 340 K for the most stable sequences (according to PFA) the weakest NN-pairs are concentrated at the six outermost base positions (at each duplex end). At 360 K (which is above the melting temperature of the duplexes) the NN pair stabilities follow a parabolic position dependence

reduced contribution to duplex free energy. At a still lower temperature of about 310 K the positional weights converge to $w_i(x) = 1$, which is equivalent to the TSNN model

Conclusion

In this paper we studied, experimentally and theoretically, the stability of short ($l < 26$ bp) linear surface-bound oligonucleotide duplexes with single base defects. We demonstrated that the rationale behind positional dependent models of oligonucleotide duplex stability is the partial denaturation of the duplexes. We have shown, that the strong influence of the defect position on mismatch discrimination [11-14, 16, 49] and the influence of the sequence context – beyond nearest neighbors [14, 34] can be quantitatively inferred from a molecular zipper model. Partial (end-domain-)denaturation of the duplex as proposed by us in [26] as well as in [16, 24, 25] results in a positional influence that is entropic in nature. The zipping process is modulated by the sequential arrangement of the base pairs. The model confirms the observed influence of the sequence context beyond the nearest-neighbors. Further the zipper model provides a theoretical foundation to the positional dependent nearest-neighbor model of Zhang et al. [17].

In the commonly employed two-state nearest-neighbor model, nucleic acid duplex hybridization/denaturation is considered to be an all-or-none process. According to literature indeed end-fraying effects are expected to be small beyond three bases [34], however, in our studied case, we conclude that end-fraying plays a non-negligible role. This is surprising since the dissociation probability of individual base pairs decreases towards the center of the duplex in an exponential fashion (see Additional file 1, Fig. S4) and remains very low for most NN-pairs.

We propose that the effect of the defect position on probe-target binding affinities becomes apparent in the hybridization signal intensities due to the unavoidable probe polydispersity of optical synthesis. It indeed appears that the positional dependence of single base MM discrimination is more commonly observed on photolithographically produced DNA oligonucleotide arrays [11-14] rather than (in large scale studies) on spotted microarrays [40, 50, 51] or in solution-phase experiments. We notice, however, that in small studies (investigating few sequences) a positional influence in solution [52] and on spotted microarrays [49] has been reported. The probe polydispersity in our experiments smoothes out the steep sigmoid relation between the hybridization intensity and binding free energy ΔG_0 that is expected for defect free probes, and explains why (within a relatively broad range of $d\Delta G^0_{D17} \approx 20$ kcal/mol) variations

of the binding free energies – like for example the influence of the defect position – are reflected (by means of an approximately linear relation) in the hybridization signal intensities

Authors' contributions

TN developed the experimental setup and methods, carried out the experiments and statistical data analysis, computational modeling and drafted the manuscript. TM participated in the development of the experimental setup and in data interpretation and helped to draft the manuscript. JK and WM participated in the modeling and helped to draft the manuscript. AO conceived of the study, and participated in its design and coordination and aided in drafting the manuscript.

Additional material

Additional file 1

Supplementary material. We provide additional figures illustrating our experimental data and theoretical analysis. We further provide the detailed analytical derivation of the defect positional influence on homopolymer sequences.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2105-9-509-S1.pdf]

References

1. Nelson JW, Martin FH and Tinoco I: **DNA and RNA Oligomer Thermodynamics – The Effect of Mismatched Bases on Double-Helix Stability.** *Biopolymers* 1981, 20(12):2509-2531.
2. Patel DJ, Kozlowski SA, Marky LA, Rice JA, Broka C, Dallas J, Itakura K and Breslauer KJ: **Structure, dynamics, and energetics of deoxyguanosine-thymidine wobble base pair formation in the self-complementary d(cgtgaattcgcg) duplex in solution.** *Biochemistry* 1982, 21(3):437-444.
3. Conner BJ, Reyes AA, Morin C, Itakura K, Teplitz RL and Wallace RB: **Detection of sickle-cell beta-s-globin allele by hybridization with synthetic oligonucleotides.** *Proceedings of the National Academy of Sciences of the United States of America* 1983, 80:278-282.
4. Devoe H and Tinoco I: **Stability of helical polynucleotides – base contributions.** *Journal of Molecular Biology* 1962, 4(5):500-8.
5. Borer PN, Dengler B, Tinoco I and Uhlenbeck OC: **Stability of Ribonucleic-Acid Double-Stranded Helices.** *Journal of Molecular Biology* 1974, 86(4):843-853.
6. Breslauer KJ, Frank R, Blocker H and Marky LA: **Predicting DNA Duplex Stability from the Base Sequence.** *Proceedings of the National Academy of Sciences of the United States of America* 1986, 83(11):3746-3750.
7. SantaLucia J and Hicks D: **The thermodynamics of DNA structural motifs.** *Annual Review of Biophysics and Biomolecular Structure* 2004, 33:415-440.
8. SantaLucia J: **A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, 95(4):1460-1465.
9. Markham NR and Zuker M: **DINAMelt web server for nucleic acid melting prediction.** *Nucleic Acids Research* 2005, 33:W577-W581.
10. Allawi HT and SantaLucia J: **Thermodynamics and NMR of internal GT mismatches in DNA.** *Biochemistry* 1997, 36(34):10581-10594.
11. Wick LM, Rouillard JM, Whittam TS, Gulari E, Tiedje JM and Hashsham SA: **On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes.** *Nucleic Acids Research* 2006, 34(3):e26.
12. Pozhitkov A, Noble PA, Domazet-Loso T, Nolte AW, Sonnenberg R, Staehler P, Beier M and Tautz D: **Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted.** *Nucleic Acids Research* 2006, 34(9):e66.
13. Naiser T, Mai T, Michel W and Ott A: **Versatile maskless microscope projection photolithography system and its application in light-directed fabrication of DNA microarrays.** *Review of Scientific Instruments* 2006, 77(6):063711.
14. Naiser T, Kayser J, Mai T, Michel W and Ott A: **Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide duplexes: comparison of single base mismatches and base bulges.** *BMC Biotechnology* 2008, 8:48.
15. Naiser T: **Characterization of Oligonucleotide Microarray Hybridization: Microarray Fabrication by Light-Directed in situ Synthesis – Development of an Automated DNA Microarray Synthesizer, Characterization of Single Base Mismatch Discrimination and the Position-Dependent Influence of Point Defects on Oligonucleotide Duplex Binding Affinities.** *PhD thesis, University of Bayreuth* 2008 http://opus.ub.uni-bayreuth.de/volltexte/2008/461/
16. Rennie C, Noyes HA, Kemp SJ, Hulme H, Brass A and Hoyle DC: **Strong position-dependent effects of sequence mismatches on signal ratios measured using long oligonucleotide microarrays.** *BMC Genomics* 2008, 9:317.
17. Zhang L, Miles MF and Aldape KD: **A model of molecular interactions on short oligonucleotide microarrays.** *Nature Biotechnology* 2003, 21(7):818-821.
18. Naef F and Magnasco MO: **Solving the riddle of the bright mismatches: Labeling and effective binding in oligonucleotide arrays.** *Physical Review E* 2003, 68:011906.
19. Mei R, Hubbell E, Bekiranov S, Mittmann M, Christians FC, Shen MM, Lu G, Fang J, Liu WM, Ryder T, Kaplan P, Kulp D and Webster TA: **Probe selection for high-density oligonucleotide arrays.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, 100(20):11237-11242.
20. Binder H, Preibisch S and Kirsten T: **Base pair interactions and hybridization isotherms of matched and mismatched oligonucleotide probes on microarrays.** *Langmuir* 2005, 21(20):9287-9302.
21. Held GA, Grinstein G and Tu Y: **Relationship between gene expression and observed intensities in DNA microarrayse modeling study.** *Nucleic Acids Research* 2006, 34(9):e70.
22. Gibbs JH and Dimarzio EA: **Statistical mechanics of helix-coil transitions in biological macromolecules.** *Journal of Chemical Physics* 1959, 30:271-282.
23. Kittel C: **Phase transition of a molecular zipper.** *American Journal of Physics* 1969, 37(9):917-&.
24. Deutsch JM, Liang S and Narayan O: **Modelling of microarray data with zippering.** *Preprint q-bio.BM/0406039 v1* 2004.
25. Binder H: **Thermodynamics of competitive surface adsorption on DNA microarrays.** *Journal of Physics-Condensed Matter* 2006, 18(18):S491-SS23.
26. Naiser T, Ehler O, Mai T, Michel W and Ott A: **Hybridization to surface-bound oligonucleotide probes: influence of point defects,** 2006.
27. Fodor SPA, Read JL, Pirrung MC, Stryer AT, Land Lu and Solas D: **Light-directed, spatially addressable parallel chemical synthesis.** *Science* 1991, 251(4995):767-773.
28. McGall GH, Barone AD, Diggelmann M, Fodor SA, Gentalen E and Ngo N: **The efficiency of light-directed synthesis of DNA arrays on glass substrates.** *Journal of the American Chemical Society* 1997, 119(22):5081-5090.
29. Singh-Gasson S, Green RD, Yue YJ, Nelson C, Blattner F, Sussman MR and Cerrina F: **Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array.** *Nature Biotechnology* 1999, 17(10):974-978.

30　Luebke KJ, Balog RP, Mittelman D and Garner HR: **Digital optical chemistry: A novel system for the rapid fabrication of custom oligonucleotide arrays.** Microfabricated Sensors, Application of Optical Technology for DNA Analysis American Chemical Society Publications Richard Kordal, Usmani, Wai Tak Law 2002, 815:87–106

31　Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg JP, Ballin J, McCormick M, Norton J, Pollock T, Sumwalt T, Butcher L, Porter D, Molla M, Hall C, Blattner F, Sussman MR, Wallace RL, Cerrina F and Green RD: **Gene expression analysis using oligonucleotide arrays produced by maskless photolithography.** Genome Research 2002, 12(11):1749–1755

32　Le Berre V, Trevisiol E, Dagkessamanskaia A, Sokol S, Caminade AM, Majoral JP, Meunier B and Francois J: **Dendrimeric coating of glass slides for sensitive DNA microarrays analysis.** Nucleic Acids Research 2003, 31(16):e88.

33　Hasan A, Stengele KP, Giegrich H, Cornwell P, Isham KR, Sachleben RA, Pfleiderer W and Foote RS: **Photolabile protecting groups for nucleosides: Synthesis and photodeprotection rates.** Tetrahedron 1997, 53(12):4247–4264

34　Hall TS, Pancoska P, Riccelli PV, Mandell K and Benight AS: **Sequence context and thermodynamic stability of a single base pair mismatch in short deoxyoligonucleotide duplexes.** Journal of the American Chemical Society 2001, 123(47):11811–11812.

35　Craig ME, Crothers DM and Doty P: **Relaxation kinetics of dimer formation by self complementary oligonucleotides.** Journal of Molecular Biology 1971, 62(2):383–8.

36　Halperin A, Buhot A and Zhulina EB: **Sensitivity, specificity, and the hybridization isotherms of DNA chips.** Biophysical Journal 2004, 86(2):718–730

37　Burden CJ, Pittelkow YE and Wilson SR: **Statistical Analysis of Adsorption Models for Oligonucleotide Microarrays.** Statistical Applications in Genetics and Molecular Biology 2004, 3:35

38　Held GA, Grinstein G and Tu Y: **Modeling of DNA microarray data by using physical properties of hybridization.** Proceedings of the National Academy of Sciences of the United States of America 2003, 100(13):7575–7580

39　Binder H, Kirsten T, Loeffler M and Stadler PF: **Sensitivity of microarray oligonucleotide probes: Variability and effect of base composition.** Journal of Physical Chemistry B 2004, 108(46):18003–18014.

40　Fish DJ, Horne MT, Brewood GP, Goodarzi JP, Alemayehu S, Bhandiwad A, Searles RP and Benight AS: **DNA multiplex hybridization on microarrays and thermodynamic stability in solution: a direct comparison.** Nucleic Acids Research 2007, 35(21):7197–7208

41　Carlon E and Heim T: **Thermodynamics of RNA/DNA hybridization in high-density oligonucleotide microarrays.** Physica A-Statistical Mechanics and its Applications 2006, 362(2):433–449

42　Peterson AW, Wolf LK and Georgiadis RM: **Hybridization of mismatched or partially matched DNA at surfaces.** Journal of the American Chemical Society 2002, 124(49):14601–14607

43　Glazer M, Fidanza JA, McGall GH, Trulson MO, Forman JE, Suseno A and Frank CW: **Kinetics of oligonucleotide hybridization to photolithographically patterned DNA arrays.** Analytical Biochemistry 2006, 358(2):225–238.

44　Vainrub A and Pettitt BM: **Coulomb blockage of hybridization in two-dimensional DNA arrays.** Physical Review E 2002, 66(4):041905

45　Halperin A, Buhot A and Zhulina EB: **Brush effects on DNA chips: thermodynamics, kinetics, and design guidelines.** Biophysical Journal 2005, 89(2):796–811

46　Jobs M, Fredriksson S, Brookes AJ and Landegren U: **Effect of oligonucleotide truncation on single-nucleotide distinction by solid-phase hybridization.** Analytical Chemistry 2002, 74:199–202.

47　Halperin A, Buhot A and Zhulina EB: **On the hybridization isotherms of DNA microarrays: the Langmuir model and its extensions.** Journal of Physics-Condensed Matter 2006, 18(18):S463–S490

48　Forman JE, Walton ID, Stern D, Rava RP and Trulson MO: **Thermodynamics of duplex formation and mismatch discrimination on photolithographically synthesized oligonucleotide arrays.** Molecular Modeling of Nucleic Acids 1998, 682:206–228

49　Dorris DR, Nguyen A, Gieser L, Lockner R, Lublinsky A, Patterson M, Touma E, Sendera TJ, Elghanian R and Mazumder A: **Oligodeoxyribonucleotide probe accessibility on a three-dimensional DNA microarray surface and the effect of hybridization time on the accuracy of expression ratios.** BMC Biotechnology 2003, 3:6.

50　Urakawa H, Noble PA, El Fantroussi S, Kelly JJ and Stahl DA: **Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses.** Applied and Environmental Microbiology 2002, 68:235–244

51　Weckx S, Carlon E, De Vuyst L and Van Hummelen P: **Thermodynamic behavior of short oligonucleotides in microarray hybridizations can be described using Gibbs free energy in a nearest-neighbor model.** Journal of Physical Chemistry B 2007, 111(48):13583–13590

52　Kierzek R, Burkard ME and Turner DH: **Thermodynamics of single mismatches in RNA duplexes.** Biochemistry 1999, 38(43):14214–14223

53　Zhu J and Wartell RM: **The effect of base sequence on the stability of RNA and DNA single base bulges.** Biochemistry 1999, 38(48):15986–15993.

Exhibit 4:   Pozhitkov et al., "Test of rRNA Hybridization to Microarrays Suggest that Hybridization Characteristics of Oligonucleotide Probes for Species Discrimination Cannot be Predicted," *Nucleic Acids Research* 34(9):e66 (2006)

Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted

Alex Pozhitkov[1,3], Peter A. Noble[1], Tomislav Domazet-Lošo[2], Arne W. Nolte[3], Rainer Sonnenberg[3], Peer Staehler[4], Markus Beier[4] and Diethard Tautz[3,*]

[1]Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA, [2]Ruder Bošković Institute, Division of Molecular Biology, Zagreb, HR-10002, Croatia, [3]Institute for Genetics, Cologne, D-50674, Germany and [4]Febit Biotech GMBH, Im Neuenheimer Feld 515, D-69120 Heidelberg, Germany

ABSTRACT

Hybridization of rRNAs to microarrays is a promising approach for prokaryotic and eukaryotic species identification. Typically, the amount of bound target is measured by fluorescent intensity and it is assumed that the signal intensity is directly related to the target concentration. Using thirteen different eukaryotic LSU rRNA target sequences and 7693 short perfect match oligonucleotide probes, we have assessed current approaches for predicting signal intensities by comparing Gibbs free energy ($\Delta G°$) calculations to experimental results. Our evaluation revealed a poor statistical relationship between predicted and actual intensities. Although signal intensities for a given target varied up to 70-fold, none of the predictors were able to fully explain this variation. Also, no combination of different free energy terms, as assessed by principal component and neural network analyses, provided a reliable predictor of hybridization efficiency. We also examined the effects of single-base pair mismatch (MM) (all possible types and positions) on signal intensities of duplexes. We found that the MM effects differ from those that were predicted from solution-based hybridizations. These results recommend against the application of probe design software tools that use thermodynamic parameters to assess probe quality for species identification. Our results imply that the thermodynamic properties of oligonucleotide hybridization are by far not yet understood.

INTRODUCTION

High throughput technologies, such as DNA microarrays, have significant potential for identifying organisms in many areas of biomedical science, including health care, biological defense and environmental monitoring. Several microarray platforms are currently used: dot blots on synthetic membranes or planar arrays (1,2) and gel-pad microarrays on glass slide (3–5). In addition, several platforms are under development: microbead microarrays (6,7) and electronic (8,9) and cantilever arrays (10). All platforms share the common attribute that a sensor detects a signal from target sequences hybridized to immobilized oligonucleotide probes. The intensity of this signal provides a measure of the amount of bound nucleic acid from a sample.

Ribosomal RNAs (rRNA) are particularly suitable for species identification procedures, because they occur universally, contain conserved as well as divergent regions, and are highly abundant in cells. Identification of microorganisms relies heavily on rRNA hybridization schemes (11,12), while applications for small eukaryotic soil or water organisms are currently emerging (14,15,18). The promise of these latter applications is that PCR amplification steps may not be required for detection, since multicellular organisms contain a sufficient amount of rRNA to allow direct detection of single individuals on a microarray platform (11–13,18).

In comparison to standard microarray applications for detecting specific mRNAs, there are extended requirements for the specific and reliable detection of organisms. First, since it is necessary to potentially distinguish closely related species, which differ only at a few nucleotide positions, one can only use relatively short oligonucleotides as probes, to ensure specificity. Second, because of the same reason, one has often only a limited set of options for choosing specific probes. And

*To whom correspondence should be addressed. Tel: 0049 221 470 2465; Fax: 0049 221 470 5975; Email: tautz@uni-koeln.de

finally, it is of particular importance that the specific probes yield a high signal to noise ratio, i.e. can discriminate accurately between perfectly matching and slightly mismatching targets.

Accordingly, it is necessary to have a reliable predictor for the hybridization performance of a specific probe. Tiling experiments with probes along specific mRNAs have shown that there can be huge differences in hybridization efficiency of probes (19,20). Furthermore, it has become clear that the simple notion that short oligonucleotides with a mismatch (MM) should hybridize less efficiently than perfect match (PM) probes is not always applicable. It has been shown that the hybridization intensity of MM probes can depend on the nucleotide type (i.e. A, C, G or T) and position of the MM relative to the termini (4,16) and that some MM probes yield higher signal intensities to the target than those of corresponding PM probes (17).

The focus of this study was to assess the utility of *in silico* predictions of probe-target duplex stabilities using DNA microarrays for detecting rRNA sequences in the context of possible applications for species identification. In particular, we investigate how well one can predict the hybridization performance of particular probes in the context of secondary structure predictions for the rRNA. In addition, we study the effects of single-base pair mismatches of all possible types and positions on probe-target hybridizations.

Our specific objectives were (i) to generate a set of probes forming a PM with target rRNA sequences, (ii) to measure the signal intensity of each probe on a microarray and to correlate fluorescent intensity values to theoretically-calculated duplex stability measures and (iii) to systematically assess the influence of single-base pair mismatches on signal intensity values of known target sequences.

We report lack of a simple relationship between hybridizations of probe-target duplexes as inferred from signal intensity values and *in silico* predictions based on Gibbs free energies. On the other hand, we can show that type and position of the MM significantly affects signal intensities of target sequences. Most interestingly, the order of stabilities of MM pairs in microarrays are different from that observed in solution, with pyrimidine–pyrimidine MM pairs being more stable than purine–purine pairs. However, even for these results the variances were high and cannot be explained for each individual oligonucleotide. Hence, it is currently not possible to predict *in silico* the performance of particular probes in microarray experiments. Accordingly we conclude that microarray designs for organism identification via rRNA hybridization will require meticulous testing of all possible oligonucleotide combinations.

MATERIALS AND METHODS

Experimental material

The ribosomal rRNA targets were derived from two different projects. For the first project, we used D3–D5 expansion segment fragments from the LSU of organisms that are present in the meiobenthos (15,18). These experiments were done in conjunction with Febit GmbH (Heidelberg), which includes also the systematic study of PM versus MM comparisons. In a second project, we have used D1–D2 expansion segment

Table 1. Sequences used and numbers of perfect match (PM) and MM probes by sequence

Sequence	Organism	Accession no	Number of probes	
			PM	MM
1	Algae	DQ086764	19	2340
2	Chironomid	DQ086592	47	2820
3	Harpacticoid	DQ086556	42	2520
4	Ostracod	DQ086565	46	2760
Cb	*Caenorhabditis briggsae*[a]	—	824	—
Ce	*Caenorhabditis elegans*[a]	—	803	—
Cr	*Caenorhabditis remanei*[a]	—	824	—
Po	*Panagrolaimus spec*[a]	—	942	—
Pm	*Plectus minimus*[a]	—	965	
Rb	*Rhabditis belari*[a]	—	779	—
Rd	*Rhabditis dolichura*[a]	—	795	—
Rt	*Rhabditis terricola*[a]	—	784	—
Tr	*Theristus rhabditidae*[a]	—	803	—
			Total 7693	10 440

[a]Species designation was done by E. Schierenberg (University of Cologne)

fragments from nine nematode species, for which we constructed PM tiling arrays in conjunction with NimbleGen Systems Inc. (Madison).

Target preparation

For the first set of experiments, cloned rDNA fragments (18) from four organisms were used (Table 1). The sequences were cloned into a pZErO-2 vector (Invitrogen Inc.). Depending on the orientation of the insert, the plasmids were cut with either SpeI or XbaI restriction enzymes and *in vitro* transcribed with SP6 or T7 RNA polymerase, respectively. The transcription and labeling mix contained 18 µl of a master-mix (10 mM ATP, CTP, GTP 8 µl each; 10 mM UTP 6 µl; 1 mM Chroma-Tide Alexa Fluor 546-14-UTP 20 µl; 10× Transcription buffer 16 µl; 40u/µl RNasin 8 µl); 2 µl of SP6 or T7 polymerase 20 u/µl; and 20 µl of the linearized plasmid at 50 ng/µl.

For the second set of experiments, ribosomal rRNA templates from nine nematode species were derived from a project in which the D1–D2 region of the LSU rRNA was sequenced (Table 1). The sequences were amplified using universal primers (28sFw-tailT3 5'-AATTAACCCTCACTAAAGGG-AGCGGAGGAAAAGAAACTA-3'; 28sRew 5'-TACTAGA-AGGTTCGATTAGTC-3') of which the forward primer carries a tail with a T3-RNA Polymerase initiation site at its 5' end. PCR products obtained with these primers were directly used for *in vitro* transcription. The transcription was performed with the MEGAscript Kit (Ambion) according to the instructions of the supplier. The master-mix was supplemented with 1.875 mM biotin-conjugated UTP (PerkinElmer) and 1.875 mM biotin-conjugated CTP (PerkinElmer) to label all transcripts.

Hybridization

Each of the rRNAs were diluted in hybridization solution (5× SSC, 0.2 mg/ml BSA, 12 mM ribonuclease inhibitor—Ribonucleoside Vanadyl Complex; New England Biolabs) to a final volume of 100 µl (3.75 ng/µl RNA) and heated to 80°C for 1 min. The following hybridization and washing protocol was used: (i) the microarrays were preheated to 70°C, (ii) the hybridization solution was added to each microarray and the

microarrays were incubated at 80°C for 1 min, (iii) a low-stringency hybridization was performed by incubating the microarrays at 45°C for 24 h, (iv) the microarrays were then washed with a low-stringency buffer (5× SSC at 20°C, 3-fold volume exchange), (v) the first image of the microarrays was recorded, (vi) the microarrays were washed with a high-stringency buffer (0.1× SSC at 20°C, 3-fold volume exchange) and (vii) a second image of the microarrays was recorded.

Hybridization on the NimbleGen platform was performed according to the protocol routinely used at NimbleGen. Briefly, each biotin-labeled rRNA target was separately hybridized to the specific compartment on the 12-well NimbleGen array (a single array with 12 compartments physically isolated from each other), such that no interference between targets was allowed. Hybridization conditions were similar to that of Febit microarray, namely 45°C, 1 M Na$^+$. After 16–20 h hybridization, the microarray was washed with non-stringent and stringent buffers and images were recorded.

Probe design

A set of oligonucleotide probes was generated using a C++ program specifically written for this study. The set consisted of PM 20mer probes that were complementary to the rRNA targets (see Target preparation section). Randomly selected 20 nt long portions of the target were considered as potential hybridization sites. In addition to the PM probes, single-MM variants were designed. The entire array of these variants made up a complete set to investigate the effects of every position of the 20mer and every type of the MM on signal intensity values. All probes were replicated four times to provide a measure of intra-microarray reproducibility. In total, 42 456 oligonucleotide probes were synthesized by the GENIOM One® instrument (Febit GmbH, Heidelberg, Germany) on the microarray as described previously (19).

The probes for the NimbleGen experiments were designed as a tiling set (1 nt shift) of perfectly matching 25 nt oligonucleotides to the rRNA sequences of the nine nematodes. In total, 7519 oligonucleotides were synthesized on the surface of the 12-well NimbleGen array (a single array with 12 compartments physically isolated from each other), each well containing the full set of oligonucleotides.

Oligonucleotide arrays

A light-activated *in situ* oligonucleotide synthesis was performed within the GENIOM instrument on the activated 3D reaction carrier, which contained a glass-silicon-glass sandwich, using a digital micromirror device (Texas Instruments). Four individually accessible microchannels (referred to as arrays) were etched into the silicon layer of the DNA processor and connected to the microfluidic system of the GENIOM instrument acting as a custom DNA synthesizer. Oligonucleotides were synthesized using standard DNA synthesis reagents and RayDite 3′ phosphoramidites, carrying a 5′-photolabile protective group (Proligo LLC; Boulder, CO, USA). Prior synthesis, the array surface was activated and enough distance between oligonucleotides was secured with a spacer to facilitate probe-target interaction and avoid probe–probe interference.

Thermodynamic calculations

The following thermodynamic parameters were calculated using different software tools: free energies of probe-target binding (ΔG°_h) and probe–probe dimerization (ΔG°_d) at 45°C were calculated using an Excel macro written by Matveeva *et al* (21); free energy of self-looping probes (ΔG°_p) at 45°C was determined by Mfold program (22). In addition, free energy of the local denaturation of the target rRNA (ΔG°_t), and the overall free energy of probe-target binding (ΔG°_{oh}) resulting from the consideration of all competing processes (i.e. ΔG°_h, ΔG°_d, ΔG°_p; see Discussion), were calculated using RNAstructure v. 4.2 |(23), set with a fixed temperature of 37°C and a probe concentration of 1 μM, and 1 M Na$^+$| All tools used the Nearest-Neighbor model.

Secondary structure prediction

The secondary structure of rRNA was determined by two alternative methods. First, the sequences were aligned to the best BLAST match from the European Ribosomal RNA Database (24), which contains an alignment of numerous LSU rRNA sequences with annotated secondary structure. Second, the rRNA targets were allowed to attain their lowest energy state. The free energies of the alternative folding were calculated using RNA folding software (RNAStructure).

Secondary structure of the targets used for hybridization with NimbleGen arrays was predicted only by energy minimization algorithm due to the lack of information about experimentally determined secondary structure of the D1–D2 expansion segment.

Data management and statistical analysis

The data were stored in a relational database created in Microsoft Access, which is available at http://faculty.washington.edu/pozhit/default.htm. The data were extracted through queries and analyzed in Microsoft Excel and SAS (Cary, N.C.). Principal component analysis (PCA) was employed to examine the distribution of the variables relative to signal intensity variables and to construct ordination plots. Pearson produce-moment correlation was used to determine the degree of association between variables. Linear regressions were used to estimate the relationship of one variable to another (25). The datasets were prepared for the ANOVA in the way that signal intensities of all duplexes where averaged using the median values of the four replicates. Median was used, as a measure of central tendency, which is less sensitive to outliers to account for possible hybridization artefacts. Median values of every probe containing a MM were normalized using the median of corresponding perfectly matched probe. These normalized values where then analyzed by three-way ANOVA using MM position, MM type and type of neighboring nucleotides (NN) as fixed factors. NN where defined as nucleotides located on the probe strand one position left and right from a MM position. Partial Eta squared (η^2) was used as a measure of the degree of association between normalized signal intensity and analyzed factors. The Hochberg's GT2 test was used for post-hoc analysis of contrast and pair-wise comparisons between means.

An artificial neural network (ANN) package (Neuroet, 26) was used to investigate the nonlinear relationships among input variables (i.e. ΔG° values) and outputs (i.e. signal
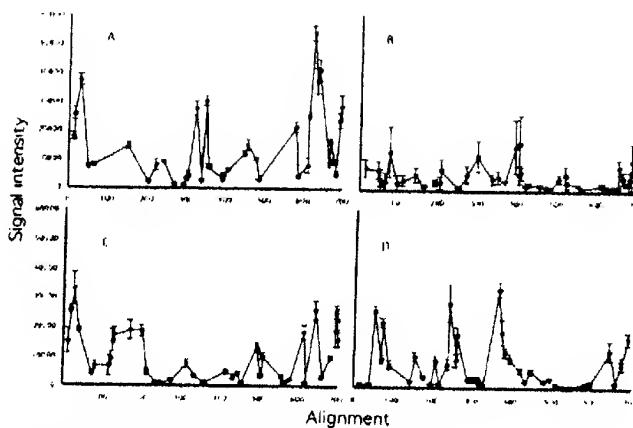
Figure 1. Signal intensity profiles of PM probes as a function of their position along target rRNAs. Error bars reflect the variance between the four replicates. The x-axis represents the position determined from the alignment based on the secondary structure predictions using the LSU database. (A) sequence 1, (B) sequence 2, (C) sequence 3 and (D) sequence 4.

intensity values). Unless otherwise specified, the following settings were used for training NNs: input and output scaling was set to standard linear (0,1); the logistic transfer function was used for hidden neurons and pure linear transfer function was used for output neurons; 80% of the data were used for training, 10% was used for testing and 10% was used for validating the NN; and, Levenberg–Marquardt error minimization was used to train the NN. The architectures of all NNs were optimized prior to conducting analyses by adjusting the number of hidden neurons (1 to 8) and identifying the architecture that provided the best predictive model. Comparison of different predictive models was conducted by computing their median Akaike's Information Criterion corrected (AICc) value (27) and determining the probability that one model was better than another. The model yielding the lowest AICc score contained the optimal number of hidden neurons.

RESULTS

In our first experiment, we constructed a set of PM probes for four different LSU rRNA fragments from meiobenthos organisms and synthesized every possible MM combination for all PM probes (Table 1). The hybridization profiles of PM probes to their respective target revealed large differences (up to 70-fold) in signal intensities by alignment position (Figure 1), similarly to what has been observed previously with mRNA targets (19,20). Matveeva et al. (19) had suggested that thermodynamic properties of probe folding and probe hybridization could partly explain these differences in hybridization efficiency. Luebke et al. (20) suggested that the predicted free energy of hybridization minus the predicted free energy for intramolecular folding of the probe provides a partial explanation, while no consistent correlation was found with the secondary structure of the mRNA targets.

Table 2 Comparison of free energies of LSU-RNA folding (kcal/mol)

Sequence	Alignment prediction	Minimum energy
1	122.6	−216.7
2	117.9	232.5
3	−126.3	−248.6
4	−119.5	250.5

Given that rRNA is known to form by far more extensive secondary structures than mRNA, we reasoned that if there would be any calculable effect of secondary structure on hybridization efficiency, it should be most pronounced for rRNA targets. Thus, in addition to calculating the parameters suggested by Matveeva et al. (19) and Luebke et al. (20), we considered also the free energy of the secondary structure of the rRNA.

Relationship of Gibbs free energy terms to signal intensity values of PM duplexes

To ensure that all possible known parameters are assessed, we calculated various Gibbs free energy terms singly or in combination using three different programs, which all consider nearest-neighbor models (see Materials and Methods). This includes the predicted free energy of hybridization (probe–target binding—$\Delta G'_b$), probe hybridization (probe–probe dimerization—$\Delta G'_d$), free energy for intramolecular folding of the probes (self-looping of probes—$\Delta G'_p$), free energy of the local denaturation of the target rRNA (ΔG°_t) and the overall free energy of probe–target binding (ΔG°_{th}) resulting from the consideration of all competing processes (i.e. $\Delta G'_b$, $\Delta G'_d$, $\Delta G''_p$, see Discussion). For considering secondary structure elements in rRNA, one can either use the secondary structure predictions inferred from alignments and experimental validation in ribosomes [taken from 'The European Ribosomal RNA

Table 4. R^2-values based on linear models for the regression between various change in Gibbs free energy terms and signal intensity values as a function of sequence for hybridization with the nematode sequences

Sequence [a]	n	Oligoarray (45°C)		RNAstructure (37°C, 1 µM probe)				
		ΔG^o_n	ΔG^o_d	ΔG^o_b	ΔG_d	$\Delta G'_b$	ΔG_r	ΔG_{ob}
Ch	824	—[b]	0.39	0.20	—	0.13	0.38	0.39
Ce	803	—	0.24	0.07	—	—	0.29	0.20
Cr	824	—	0.34	0.17	—	0.09	0.34	0.27
Pm	965	—	0.19	0.19	0.12	0.19	0.40	0.40
Po	932	0.08	0.49	0.40	0.66	0.05	—	—
Rb	779	0.67	—	0.11	—	0.06	0.15	0.21
Rd	795	0.74	—	0.08	0.06	—	0.09	0.07
Rr	784	0.54	—	0.10	0.16	—	0.06	—
Tr	803	—	0.16	0.14	0.05	—	0.16	0.07

Secondary structure of the target was determined by energy minimization using RNAstructure
[a]Refer to Table 1 for sequences.
[b]Not significant at $\alpha = 0.05$

Table 5. Correlation coefficients of variables relative to PCA axes by sequence (based on RNAStructure)

Probe characteristics	Sequence											
	1 PC axes (n = 156)			2 PC axes (n = 188)			3 PC axes (n = 168)			4 PC axes (n = 184)		
	1	2	3	1	2	3	1	2	3	1	2	3
GC	−0.60	−0.75	—	−0.50	−0.77	0.30	−0.62	−0.62	—	−0.54	0.56	−0.58
ΔG^o_n	0.53	0.77	0.19	0.41	0.71	0.49	0.66	0.65	0.26	0.54	0.75	0.33
ΔG_d	0.57	0.31	0.40	0.30	0.51	0.70	0.19	0.19	0.65	0.41	0.37	0.41
ΔG^o_b	0.55	0.40	—	—	0.75	0.38	0.17	—	0.77	—	0.49	0.52
ΔG^o_r (aligned)	0.78	0.25	0.47	0.75	0.30	0.46	0.62	0.53	—	0.83	—	0.44
ΔG_{ob} (aligned)	0.59	0.68	−0.36	0.92	—	—	0.86	—	—	0.67	0.33	0.58
ΔG^o_r (min energy)	0.79	—	−0.47	−0.59	0.70	—	−0.33	0.84	—	0.94	—	—
ΔG_{ob} (min energy)	−0.52	0.61	0.58	0.80	−0.38	−0.24	0.68	−0.52	—	−0.84	−0.33	—
Ln signal intensity (stringent)[a]	0.50	−0.37	0.36	−0.37	—	0.64	−0.36	—	0.80	—	0.59	0.58
Proportion of eigenvalues	0.39	0.24	0.15	0.34	0.29	0.19	0.33	0.26	0.20	0.39	0.20	0.20

'—' Not significant at $\alpha = 0.05$
[a]Similar results were obtained for non-stringent conditions and not shown for brevity

point on the plots by its corresponding signal intensity value. Examination of the four ordination plots revealed no obvious relationship between any of the variables and signal intensity values (data not shown). To more thoroughly investigate the relationship between ΔG^o terms and signal intensity values, signal intensity values were included as a variable in PCA. PCA results of the data from different target sequences revealed that 78–82% of the total matrix variance was explained by three principal axes, with PC1 explaining 33–39%, PC2 explaining 20–29%, and PC3 explaining 15–20% of the total matrix variance (Table 5). However, Pearson correlation coefficients of the variables relative to the PC axes revealed inconsistent results for the data from different target sequences. For example, in the case of sequences 1 and 4, PC1 was most strongly positively correlated to $\Delta G'_r$, while sequences 2 and 3 PC1 was negatively correlated to $\Delta G'_r$. For sequences 2 and 3, ΔG^o_{ob} was most strongly correlated to PC1, while this was negatively correlated for sequences 1 and 4. Similar results were also obtained for the other PC axes, indicating differences in the ordination of variables for data from different target sequences, which was also evident in the two dimension plots (data not shown).

The same analysis was carried out for the second experiment on the NimbleGen arrays. Similarly, examination of the nine ordination plots revealed no obvious relationship between any of the variables and signal intensity values

(data not shown). In order to more thoroughly investigate the relationship between ΔG^o terms and signal intensity values, the signal intensity values were included as a variable in PCA. PCA results of the data from different target sequences revealed that 83–91% of the total matrix variance was explained by three principal axes, with PC1 explaining 36–58%, PC2 explaining 17–29%, and PC3 explaining 14–24% of the total matrix variance (Supplementary Tables S1–S3). However, Pearson correlation coefficients of the variables relative to the PC axes revealed inconsistent results for the data from different target sequences.

To assess hidden nonlinear relationships, ANN analysis was used to investigate the relationship between ΔG^o terms and signal intensity values, because neural networks have been shown to handle noisy, nonlinear data better than conventional linear approaches, such as PCA (28). For these analyses, the optimal number of hidden neurons was found to be 4, when ΔG terms are used as inputs and signal intensity values are used as outputs. A model of the relationship between ΔG^o terms and signal intensity values was generated by training an ANN using the data from one target sequence and cross validating the generated model by using data from another target sequence.

The correlation coefficients between actual and predicted signal intensity values of the models are shown in Table 6. A correlation close to 1 or −1 indicates that a model accurately

Table 6. Cross-validation (CV) of ANN results. ΔG° terms from RNA structure were used as inputs and signal intensity values were used as outputs

ANN trained by target sequence	n	Correlation coefficient by target sequence used for CV			
		1	2	3	4
1	156	0.94	0.34	—	0.29
2	188	0.25	0.82	—	—
3	167	0.30	0.26	0.98	0.25
4	184	0.37	—	—	0.98

In all cases, the ANN were trained using four hidden neurons. 10% of the data were used for testing, and another 10% was used for validation of the model. '—' Not significant at α = 0.05.

Table 7. Three-way ANOVA of normalized signal intensities obtained in the stringent hybridization experiment

Component of variance	df	F	P	Partial η²
MM position (MMP)	17	45.86701	1.20E−146	0.096
MM type (MMT)	11	41.68068	1.69E−88	0.059
NN	4	33.89908	4.52E−28	0.018
MMP × MMT	187	1.38596	0.000454	0.034
MMP × NN	34	2.864812	5.75E−08	0.013
MMT × NN	44	2.036145	6.41E−05	0.012
MMP × MMT × NN	374	0.831392	ns	0.012
Error	7322			

predicts signal intensity values when provided with ΔG° terms, while a correlation value close to zero implies no correlation. We also included the correlation coefficient for each ANN trained with the same data since it represents the 'best' possible correlation for each model. Note that the 'best' possible correlations were based on the analysis of all predicted and actual signal intensity values in the data from one sequence. The reason why the 'best' correlations were not exactly 1 or −1 was because only 80% of the data were used to train the ANN model. The remaining (20%) of the data were used for local testing and validation of the model.

Poor correlations of ANN predictions to actual values could be attributed to over- or under-training of the ANNs. For example, an over-trained ANN learns to memorize the training data, and consequently generates high correlations between predicted and actual values for data it was trained on, but poor or no correlations for test data that was not used for training. We carefully trained each ANN model to generalize predictions by optimizing the architecture of the model prior to training, and by stopping training when there was no change in the error over a specified period of time or, after a specified number of iterations [see ref. (26)]. This approach ensured that each ANN model produced outputs that accurately predict signal intensity values for ΔG° terms not used for training. We conclude that the reason the ANN models are unable to accurately predict signal intensity values when provided with data from sequences not used for training, was because there is a poor relationship between ΔG° terms and signal intensity values. These findings corroborate the PCA results and suggest that no combinations of the ΔG° terms are major determinants for predicting signal intensity values.

An assessment of the effects of mismatches on signal intensity values

Three-way ANOVA was used to assess the effects of MM position, MM type, and the type of NN that flank a MM, on normalized signal intensity values (see Materials and Methods). The model revealed that all three factors had low, albeit significant effects on the normalized signal intensity values (Table 7). Most of the variance of normalized signal intensity was explained by MM position (9.6%), followed by MM type (5.9%), whereas NN type had the lowest effect on the observed variance among the factors (1.8%) measured by partial η². In addition, there were interactions among all combinations of two factors (Table 7). The strongest interaction was observed for MM positions and MM types (3.4%), while

interactions between MM position and NN type (1.3%) and between MM type and NN type (1.2%) were comparable. We were not able to detect significant effects of simultaneous interactions among all three factors (Table 7).

Moving the position of the MM away from the 5′ or 3′ termini to the center of the probe significantly decreased signal intensities (Figure 3). ANOVA post-hoc contrasts between means showed that duplexes with MM between positions 6 and 15 formed a homogenous group (α = 0.05) with the most pronounced effects on duplex stability. This finding indicates that the most optimal discrimination of MM from PM duplexes is provided with the MM in the middle of the duplex. However, we emphasize that this was an average result, and note that in some individual cases, MM probes with central mismatches (positions 9–11) were observed to have signal intensities that were equal or up to 1.6 times higher than that of corresponding PM probes.

A heat map on the effects of the MM type by position is shown on Figure 4. Clearly, there are differences in average signal intensity by MM type and position. Post-hoc ANOVA contrasts were able to discriminate five homogenous groups (Figure 5), two groups with clearly separated extremes (i) GA and GG mismatches (which destabilize duplexes the most) and (ii) TC, TU and TG mismatches (which destabilize duplexes the least). Differences in signal intensity values as a function of position are clearly visible for these two groups in Figure 4. These findings indicate that distinguishing PM duplexes from those containing a single-MM was highly dependent on the type of MM pairs.

To more fully understand simple patterns of MMs as a function of type and position, we pooled MM types to three categories: purine–purine, pyrimidine–pyrimidine and purine–pyrimidine MM pairs. Figure 6 shows that signal intensities of duplexes with pyrimidine–pyrimidine MM pairs were more similar to PM duplexes than purine–pyrimidine or purine–purine MM pairs. An interaction was evident at the termini of probes where differences in the normalized signal intensities among MM pairs were more pronounced towards the 3′ end of the probe. Differences in intensity values at the 5′ and 3′ end might be due to the orientation of the probe on the microarray since the 3′ end was closest to the microarray surface.

Figure 7 illustrates the effects of the type of NN that flank a MM on normalized signal intensity values. We analyzed separately the cases when a MM is located at the termini of sequence from the cases when it is located elsewhere. The reason for this is that a MM at the termini could have only one neighboring nucleotide, while in all other positions it has two neighbors. We assessed the effect of NN by categorizing
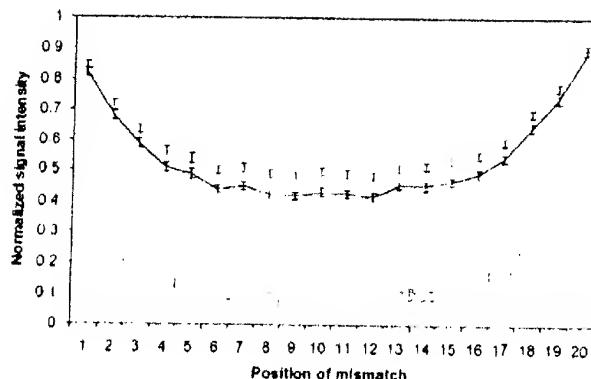
Figure 3. Average signal intensity values of MM duplexes at positions 1 to 20 normalized to that of the PM duplex (based on around 400 values per position). Error bars represent ±1 standard error of mean. Intensities from low-stringency (gray line) and high-stringency (black line) experiments are shown. Differences between low-stringency and high-stringency experiments are significant for overall dataset and for the individual MM positions at least on the $\alpha = 0.05$ level by paired t-test. Shaded cells in each row represent a homogenous set of means revealed by GT2 post-hoc analysis at $\alpha = 0.05$ level.

Figure 4. Heat map of MM type by position as a function of average signal intensity, normalized to the signal intensity of the PM duplex. Each box represents at least 120 replicates.

probes with terminal MMs into two states, those with a purine and those with a pyrimidine neighbor. Elsewhere we grouped MMs having nucleotides flanking a MM into three categories: purines only, pyrimidines only, and purine–pyrimidine combinations. In addition to the asymmetric impact of MM type at the end of the probe described earlier, we detected asymmetry at the ends of the probe concerning NN type. Figure 7 shows that purine neighbor at the 5′ end stabilized the duplex more than pyrimidine (GT2 post-hoc test, $P = 0.001$). Surprisingly, at the 3′ end the opposite trend is true—although it is not statistically significant. When non-terminus mismatches were considered, the most stabilizing effect on the duplex occurred with purine flanking neighbors. Pyrimidine flanking neighbors yielded the lowest duplex stability. Purine–pyrimidine neighbors were in the middle of these two extremes (Figure 7B, all differences are significant at $P = 0.001$ by GT2 post-hoc test). Interactions between NN type and MM position and type are significant as previously stated. However, due to

the minor effects on the variance and peculiar patterns of interaction, we excluded it from further discussion.

DISCUSSION

The thermodynamic properties of nucleic acid duplex formation and dissociation in solution have been well established (29). For example, the behavior of a probe and a target sequence in solution can be predicted by using a nearest-neighbor model (30). However, duplex formation using surface-immobilized DNA oligonucleotides is less well understood, presumably due to the complex factors affecting the kinetics and thermodynamics of target capture. Some factors affecting duplex formation on DNA microarrays include: probe density, microarray surface composition and the stabilities of oligonucleotide-target duplexes, intra- and intermolecular self-structures and RNA secondary structures
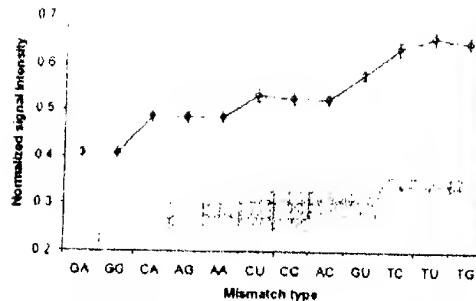
Figure 5. Average signal intensity values of MM duplexes categorized by MM type. Shaded cells in each row represent a homogenous set of means revealed by GT2 post-hoc analysis at $\alpha = 0.05$ level. Error bars represent ±1 standard error of mean. Note that each member of mirrored MM pairs (GU and TG, TC and CU, GA and AG, CA and AC) belongs to the different homogenous group. All differences within mirrored pair of mismatches are significant at least at $\alpha = 0.01$ level in GT2 pair-wise comparisons.
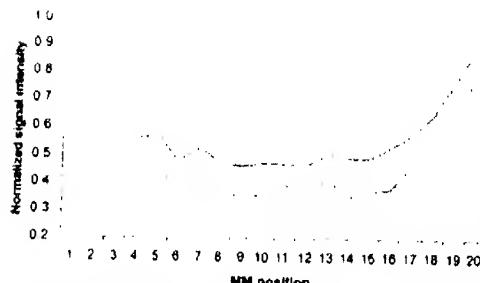
Figure 6. Average signal intensity values of MM duplexes at positions 1 to 20 normalized to signal intensity of the PM duplex by MM type. Pyrimidine pyrimidine MMs, yellow. Purine pyrimidine, red. Purine purine MMs, blue.

Nucleotide type adjacent to a mismatch

Figure 7. Average signal intensity values of MM duplexes categorized by nucleotide type flanking a MM in the probe sequence. Shaded cells in each row represent a homogenous set of means revealed by GT2 post-hoc analysis at $\alpha = 0.05$ level. Error bars represent ± standard error of mean. Upper panel: Effect of nucleotide types adjacent to a MM at the end of probe. Note that at 5' end neighboring purine residues stabilize duplex more than pyrimidine residues ($P = 0.001$), while pattern at the 3' is opposite and there is no significant difference in GT2 pair-wise comparison. Lower panel: Effect of nucleotide types flanking a MM. Purine residues are stabilizing duplex more than purine–pyrimidine combinations or pyrimidine alone. All differences are significant at $\alpha = 0.001$ level in GT2 pair-wise comparisons.

(21,31,32). We reasoned that examination of the thermodynamic stabilities of probe-target duplexes using existing models might provide valuable information on the relationships between predicted stabilities of targets hybridized to immobilized probes and their corresponding signal intensity values on DNA microarrays. We also reasoned that the position and type of MM, and the nature of neighboring bases to the MM should also affect signal intensity values.

Relationship between thermodynamic predictions and signal intensity values

As proposed by Matveeva *et al* (21), hybridizations of a target to probes on a planar microarray are affected by several overlapping processes which include, (i) the affinity of a target to bind to a probe (ΔG°_{b}), (ii) the formation of stem–loop structures of a probe (ΔG°_{p}), (iii) the formation of secondary structure (loops and helices) of a target (ΔG°_{t}) and (iv) probe to probe dimerization (ΔG°_{d}) (Figure 9). In addition, the overall

Gibbs free energy of binding (ΔG°_{Ob}) can be calculated by considering the combined effects of all four terms (i.e. ΔG°_{b}, ΔG°_{p}, ΔG°_{d} and ΔG°_{t}) on hybridization predictions (23).

ΔG°_{t} values could have been considered of special relevance for rRNA, because of the known potential to form extensive secondary structures. The values were calculated by considering the secondary structure of the targets as determined from the LSU rRNA database. Two different approaches were used to calculate ΔG°_{t} since we did not know if aligned (constrained) or not aligned (free form) secondary structure significantly affected free energies determination. The aligned folding preserves the annotated single strands while the not aligned folding allows the molecule to reach a conformation that corresponds to the calculated global energy minimum.

In our analysis, all Gibbs free energy terms were poorly correlated and linear and nonlinear regressions had low R^2-values, to signal intensity values of PM probe-target duplexes. Moreover, there does not appear to be a consistent pattern in Gibbs free energy terms by target sequence
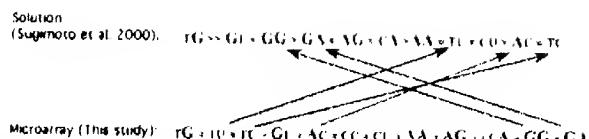
Figure 8. The order of stability of RNA/DNA duplexes with a single-base pair MM pairs in solution Sugimoto *et al.* (39) and on the microarray. For each MM pair probe DNA is on the left and target RNA is on the right. The size of the letter distinguishes purines (large) from pyrimidines (small). Lines depict major differences in the order of stability.

Furthermore, while PCA and ANN analyses were able to establish significant correlations between Gibbs free energy terms and signal intensity values when each sequence was separately analyzed, cross validation using different target sequences revealed inconsistent results. These findings indicate that Gibbs free energy terms and signal intensity values are target dependent and suggest that other factors, such as surface density of the probes (31) and/or brush effects (32), might have greater effects on signal intensity values than previously anticipated.

Thermodynamic stabilities of target RNA hybridized to immobilized oligonucleotide probes have been investigated in the following studies: (i) Naef and Magnasco (17) and Mei *et al.* (33) both described an *ad hoc* model that examined the affinity of a probe to a target based on the sum of position-dependent base-specific contributions, (ii) Zhang *et al.* (34) described an *ad hoc* model that considered position-dependent nearest-neighbor effects, (iii) Held *et al.* (35) examined the effects of free energies of RNA/DNA duplex formation and (iv) Wu and Irizarry (36) developed a model that considered both stochastic and deterministic aspects of probe-target hybridizations. The unifying features of these studies are (i) they are all based on the analysis of multiple probes targeting mRNA transcripts (i.e. expression data), (ii) with exception of Held *et al.* (35), they only considered single-base pair mismatches that occurred in the middle of the duplex (position 13 of 25mers), (iii) they assumed that binding of various RNA targets was independent and noncompetitive. Unfortunately, none of the studies satisfactorily predicted signal intensity values on oligonucleotide microarrays since there were significant disagreements between actual and predicted values.

The effect of single-base pair mismatches on duplex signal intensity values

In solution, single-base-mismatches in oligonucleotide probes can stabilize or destabilize a duplex depending on the identity of the MM, its position in the helix and its neighboring base pairs (37). Although it has been established that there are differences in experimental results conducted in solution versus those using microarrays (17), we investigated MM type and position, and neighboring base pairs on signal intensity values because the effects of these variables on planar microarrays are not well understood.

We found that the position of the MM affected duplex stability (as inferred by signal intensity values). This finding is consistent with previous studies (16) showing that terminal mismatches are less destabilizing than internal ones. We also found asymmetry in the pattern of signal intensity values by position. Specifically, normalized signal intensities among
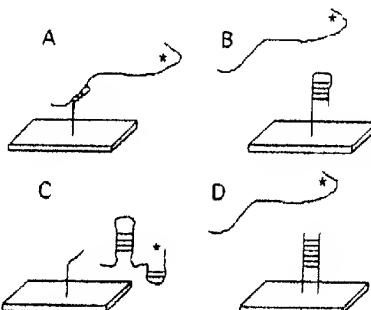
Figure 9. Depiction of four competitive processes on signal intensity values. Each panel shows a labeled (*) target and an immobilized probe on a microarray. (A) hybridization of a target to a probe, (B) probe self-folding, (C) folding of the target and (D) dimerization of adjacent probes.

MM pairs were more pronounced towards the 3' end of the probe. This phenomenon was presumably due to orientation of the probe on the microarray since the 3' end was tethered to the microarray surface. Since electrostatic effects of the microarray surface are distance dependent, mismatches closest to the 3' end might be responsible for the observed effect (38)—although further studies are needed to verify this.

Studies conducted in solution have shown that different MM types cause diverse effects on duplex stability (39). We found that the order of stabilities of MM pairs in solution were different from that observed in microarrays (Figure 8). In general, the microarray results revealed that pyrimidine-pyrimidine MM pairs were more stable (left side of Figure 8) than purine-purine MM pairs (right side of Figure 8). This result was anticipated since purines are composed of large double-ringed nucleotides that distort the geometry of the double helix—incurring a large steric and stacking cost. Hence, MM pairs containing purine destabilize the duplex and have lower signal intensity values than its corresponding PM duplex. Pyrimidine-pyrimidine mismatches, on the other hand, are composed of small single rings that do not distort the geometry of the double helix, resulting in higher stabilities and signal intensity values than MM pairs containing one or two purines. Possible reasons for the discrepancy in the order of stabilities in solution versus those in microarrays include: the number of samples examined [Sugimoto *et al.* (39) versus this study, 52 versus 10 440 MM pairs, respectively], the size of the oligonucleotide probes on the microarrays (9mers versus 20mers, respectively), and neighboring bases employed (C-MM-G, G-MM-C, C-MM-C,

G-MM-G versus every possible neighboring combination, respectively)

Interestingly, we also found some asymmetries in signal intensity values of mismatches that contain the same pair of bases (e.g. GA and AG; Figures 4 and 5) but differ only in the sense that MM nucleotide is either on the probe or target strand of the duplex. Sugimoto *et al.* (39) also found this asymmetry for mismatches occurring in short oligonucleotides in solution. This effect can currently not be explained.

The bases neighboring a probe MM can also significantly affect signal intensity values. Bases neighboring a MM at the 5'-terminus had contrasting affects on signal intensity values to those at the 3'-terminus. For example, at the 5'-terminus, purine neighbors had higher signal intensity values than pyrimidine neighbors, while at the 3'-terminus, purine neighbors had the opposite effects on signal intensity values (Figure 7). These differences may be due to steric effects of MMs at the 3'-terminus, which are close to the microarray surface. In contrast to bases neighboring a MM at the terminus, bases neighboring an internal MM yielded a consistent trend: mismatches flanked by purine neighbors had a more stabilizing effect on duplexes than other combinations. These findings are consistent with Sugimoto *et al.* (39), which showed that both the MM type and the neighboring bases of the probe influenced duplex stability.

CONCLUSION

In summary, there is little evidence to support the notion that thermodynamic parameters accurately predict signal intensity values of duplexes with rRNAs on oligonucleotide (20–25 nt) DNA microarrays. As a consequence, we recommend that thermodynamic criteria (e.g. 21, 40) not be used for designing oligonucleotide probes for species identification—instead, an empirical verification of each probe is advised to obtain the best signal intensities. Thorough empirical calibration of microarrays has recently been shown to be useful in a related field [methylation pattern analysis via microarray-based genotyping, (41)] to select best probes within one or two optimization and selection cycles. With respect to MM effects, we find that the position and type of single-base pair MM and composition of neighboring bases affected the stability of duplexes on DNA microarrays—but in different ways from what is known from experiments conducted in solution. Key differences are: (i) positional affects of MMs were asymmetric, presumably due to steric affects of mismatches close to the surface of the microarray; (ii) pyrimidine–pyrimidine MM pairs were more stable than purine–purine MM pairs and (iii) duplexes with mismatches flanked by purine neighbors were more stable than other combinations of neighbors. However, we point out that even these effects, although consistent, have only a partial predictive value.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. DeSantis,T.Z., Stone,C.E., Murray,S.R., Moberg,J.P. and Andersen,G.L. (2005) Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. *FEMS Microbiol. Lett.*, 245, 271–278.
2. Wilson,W.J., Strout,C.L., DeSantis,T.Z., Stilwell,J.L., Carrano,A.V. and Andersen,G.L. (2002) Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol. Cell. Probes* 16, 119–127.
3. Pozhitkov,A., Smidt,H., Konneke,M., Chernov,B., Yershov,G. and Noble,P.A. (2005) Evaluation of gel-pad oligonucleotide microarray technology using artificial neural networks. *Appl. Environ. Microbiol.* 71, 8663–8676.
4. Urakawa,H., Noble,P., A., El Fantroussi,S., Kelly,J.J. and Stahl,D.A. (2002) Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. *Appl. Environ. Microbiol.*, 68, 235–244.
5. Liu,W.T., Mirzabekov,A.D. and Stahl,D.A. (2001) Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. *Environ. Microbiol.* 3, 619–629.
6. Brenner,S., Johnson,M., Bridgham,J., Golda,G., Lloyd,D.H., Johnson,D., Luo,S.J., McCurdy,S., Foy,M., Ewan,M. *et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.*, 18, 630–634.
7. Brenner,S., Williams,S.R., Vermaas,E.H., Storck,T., Moon,K., McCollum,C., Mao,J.I., Luo,S.J., Kirchner,J.J., Eletr,S. *et al.* (2000) In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl Acad. Sci. USA*, 97, 1665–1670.
8. Barlaan,E.A., Sugimori,M., Furukawa,S. and Takeuchi,K. (2005) Electronic microarray analysis of 16S rDNA amplicons for bacterial detection. *J. Biotechnol.*, 115, 11–21.
9. Zimmermann,K., Eiler,T. and Scheißinger,F. (2003) Consecutive analysis of bacterial PCR samples on a single electronic microarray. *J. Microbiol. Methods*, 55, 471–474.
10. McKendry,R., Zhang,J.Y., Arntz,Y., Strunz,T., Hegner,M., Lang,H.P., Baller,M.K., Certa,U., Meyer,E., Gunderodt,H.J. *et al.* (2002) Multiple label-free biodetection and quantitative DNA-binding assays on a nanomechanical cantilever array. *Proc. Natl Acad. Sci. USA*, 99, 9783–9788.
11. Peplies,J., Lau,S.C.K., Pernthaler,J., Amann,R. and Glockner,F.O. (2004) Application and validation of DNA microarrays for the 16S rRNA-based analysis of marine bacterioplankton. *Environ. Microbiol.* 6, 638–645.
12. Small,J., Call,D.R., Brockman,F.J., Straub,T.M. and Chandler,D.P. (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl. Environ. Microbiol.*, 67, 4708–4716.
13. Chandler,D.P., Newton,G.J., Small,J.A. and Daly,D.S. (2003) Sequence versus structure for the direct detection of 16S rRNA on planar oligonucleotide microarrays. *Appl. Environ. Microbiol.*, 69, 2950–2958.
14. Blaxter,M., Elsworth,B. and Daub,J. (2004) DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.*, 271, S189–S192.
15. Markmann,M. (2000) Entwicklung und Anwendung einer 28S rDNA Sequenzdatenbank zur Ausschlusselung der Artenvielfalt limnischer Meiobenthosfauna im Hinblick auf den Einsatz moderner Chiptechnologie. PhD Thesis, Ludwig Maximilians University, Munich.

16. Urakawa,H., El Fantroussi,S., Smidt,H., Smoot,J.C., Tribou,E.H., Kelly,J.J. Noble P.A. and Stahl,D.A. (2003) Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays. *Appl. Environ. Microbiol.* 69, 2848–2856.

17. Naef,F. and Magnasco,M.O. (2003) Solving the riddle of the bright mismatches, labeling and effective binding in oligonucleotide arrays. *Phys. Rev. E Stat. Nonlin. Soft matter Phys.*, 68. Art. No. 011906 Part 1.

18. Markmann,M. and Tautz,D. (2005) Reverse taxonomy: an approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. *Phil. Trans. R. Soc. B*, 360, 1917–1926.

19. Baum,M., Bielau,S., Rittner,N., Schmid,K., Eggelbusch,K., Dahms,M., Schlauersbach,A., Tahedl,H., Beier,M., Guimil,R. *et al.* (2003) Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. *Nucleic Acids Res.*, 31, e151.

20. Luebke,K.J., Balog,R.P. and Garner,H.R. (2003) Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to RNA transcripts. *Nucleic Acids Res.*, 31, 750–758.

21. Matveeva,O.V., Shabalina,S.A., Nemtsov,V.A., Tsodikov,A.D., Gesteland,R.F. and Atkins,J.F. (2003) Thermodynamic calculations and statistical correlations for oligoprobes design. *Nucleic Acids Res.*, 31, 4211–4217.

22. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31, 3406–3415.

23. Mathews,D.H., Burkard,M.F., Freier,S.M., Wyatt,J.R. and Turner,D.H. (1999) Predicting oligonucleotide affinity to nucleic acid targets. *RNA Publ. RNA Soc.*, 5, 1458–1469.

24. Wuyts,J., Perriere,G. and de Peer,Y.V. (2004) The European ribosomal RNA database. *Nucleic Acids Res.*, 32, D101–D103.

25. Sokal,R.R. and Rohlf,F.J. (1981) *Biometry* 2nd edn. W.H. Freeman and Co., NY.

26. Noble,P.A. and Tribou,E. (2006) Neuroet: an easy-to-use artificial neural network for ecological and biological modeling. *Ecological Modelling* (in press).

27. Motulsky,H. and Christopoulos,A. (2004) *Fitting models to biological data using linear and nonlinear regression a practical guide to curve fitting.* Oxford University Press, NY.

28. Noble,P.A., Almeida,J.S. and Lovell,C.R. (2000) Application of neural computing methods for interpreting phospholipid fatty acid profiles of natural microbial communities. *Appl. Environ. Microbiol.*, 66, 694–699.

29. Marky,L.A. and Breslauer,K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, 26, 1601–1620.

30. SantaLucia,J., Allawi,H.T. and Seneviratne,A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, 35, 3555–3562.

31. Peterson,A.W., Heaton,R.J. and Georgiadis,R.M. (2001) The effect of surface probe density on DNA hybridization *Nucleic Acids Res.*, 29, 5163–5168.

32. Halperin,A., Buhot,A. and Zhulina,E.B. (2005) Brush effects on DNA chips: thermodynamics, kinetics, and design guidelines. *Biophys. J.*, 89, 796–811.

33. Mei,R., Hubbell,E., Bekiranov,S., Mittmann,M., Christians,F.C., Shen,M.M., Lu,G., Fang,J., Liu,W.M., Ryder,T. *et al.* (2003) Probe selection for high-density oligonucleotide arrays. *Proc. Natl Acad. Sci. USA*, 100, 11237–11242.

34. Zhang,L., Miles,M.F. and Aldape,K.D. (2003) A model of molecular interactions on short oligonucleotide microarrays. *Nat. Biotechnol.*, 21, 818–821.

35. Held,G.A., Grinstein,G. and Tu,Y. (2003) Modeling of DNA microarray data by using physical properties of hybridization. *Proc. Natl Acad. Sci. USA*, 100, 7575–7580.

36. Wu,Z. and Irizarry,R.A. (2004) Stochastic models inspired by hybridization theory for short oligonucleotide microarrays. In *Proceedings of RECOMB 2004.* San Diego, CA.

37. Kierzek,R., Burkard,M.E. and Turner,D.H. (1999) Thermodynamics of single mismatches in RNA duplexes. *Biochemistry*, 38, 14214–14223.

38. Vainrub,A. and Pettitt,B.M. (2002) Coulomb blockage of hybridization in two-dimensional DNA arrays. *Phys. Rev. E*, 66. Art. No. 041905 Part 1.

39. Sugimoto,N., Nakano,M. and Nakano,S. (2000) Thermodynamics-structure relationship of single mismatches in RNA/DNA duplexes. *Biochemistry*, 39, 11270–11281.

40. Tanaka,F., Kameda,A., Yamamoto,M. and Ohuchi,A. (2005) Design of nucleic acid sequences for DNA computing based on a thermodynamic approach. *Nucleic Acids Res.*, 33, 903–911.

41. Mund,C., Beier,V., Bewerunge,P., Dahms,M., Lyko,F. and Hoheisel,J.D. (2005) Array-based analysis of genomic DNA methylation patterns of the tumour suppressor gene p16(INK4A) promoter in colon carcinoma cell lines. *Nucleic Acids Res.*, 33, e73

Exhibit 5:   ClustalW2 pairwise sequence alignment results

All Databases

Enter Text Here

Help
General Help
Formats
Gaps
Matrix
References
ClustalW2 Help
ClustalW2 FAQ
Jalview Help
Scores Table
Alignment
Guide Tree
Colours
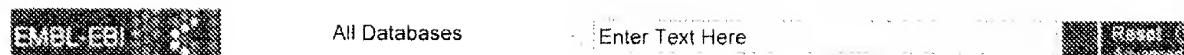
ClustalW2 Results

	Results of search
Number of sequences	2
Alignment score	39
Sequence format	Pearson
Sequence type	nt
JalView	Start Jalview
Output file	clustalw2-20090109-16515619.output
Alignment file	clustalw2-20090109-16515619.aln
Guide tree file	clustalw2-20090109-16515619.dnd
Your input file	clustalw2-20090109-16515619.input

SUBMIT ANOTHER JOB

To save a result file right-click the file link in the above table and choose "Save Target As".
If you cannot see the JalView button, reload the page and check your browser settings to enable Java Apple

Scores Table

View Output File

```
SeqA Name   Len(nt)  SeqB Name   Len(nt)  Score
================================================
1    Seq1   10       2    Seq2   10       90
================================================
```

PLEASE NOTE: Some scores may be missing from the above table if the alignment was done using multiple

View Output File

Alignment

Show Colors   View Alignment File

```
CLUSTAL 2.0.10 multiple sequence alignment

Seq1        ATTCTTGTTA- 10
Seq2        -TTCTTGTTAA 10
            ********
```

PLEASE NOTE. Showing colors on large alignments is slow

[Show Colors]   [View Alignment File]

Guide Tree

[Show as Phylogram Tree]   [Show Distances]   [View DND File]

(Seq1:0.05,Seq2:0.05);

Cladogram

[Show as Phylogram Tree]   [Show Distances]   [View DND File]

Right-click on the above tree to see display options.
Problems printing? Read how to print a Phylogram or Cladogram

Terms of Use   EBI Funding   Contact EBI   © European Bioinformatics Institute 2006-2009. EBI is an Outstation of the European Mole

Exhibit 6:   Gerry et al., "Universal DNA Microarray Method for Multiplex Detection of Low Abundance Point Mutations," *J. Mol. Biol.* 292:251–62 (1999)

JMB

Universal DNA Microarray Method for Multiplex Detection of Low Abundance Point Mutations

Norman P. Gerry[1], Nancy E. Witowski[2], Joseph Day[1], Robert P. Hammer[3], George Barany[2] and Francis Barany[1]*

[1]Department of Microbiology Hearst Microbiology Research Center, and Strang Cancer Prevention Center, Joan and Sanford I. Weill Medical College of Cornell University 1300 York Ave., Box 62, New York, NY 10021, USA

[2]Departments of Chemistry and Laboratory Medicine & Pathology, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, MN 55455, USA

[3]Department of Chemistry Louisiana State University 232 Choppin Hall, Baton Rouge LA 70803, USA

Cancers arise from the accumulation of multiple mutations in genes regulating cellular growth and differentiation. Identification of such mutations in numerous genes represents a significant challenge in genetic analysis, particularly when the majority of DNA in a tumor sample is from wild-type stroma. To overcome these difficulties, we have developed a new type of DNA microchip that combines polymerase chain reaction/ligase detection reaction (PCR/LDR) with "zip-code" hybridization. Suitably designed allele-specific LDR primers become covalently ligated to adjacent fluorescently labeled primers if and only if a mutation is present. The allele-specific LDR primers contain on their 5'-ends "zip-code complements" that are used to direct LDR products to specific zip-code addresses attached covalently to a three-dimensional gel-matrix array. Since zip-codes have no homology to either the target sequence or to other sequences in the genome, false signals due to mismatch hybridizations are not detected. The zip-code sequences remain constant and their complements can be appended to any set of LDR primers, making our zip-code arrays universal. Using the K-ras gene as a model system, multiplex PCR/LDR followed by hybridization to prototype 3 × 3 zip-code arrays correctly identified all mutations in tumor and cell line DNA. Mutations present at less than one per cent of the wild-type DNA level could be distinguished. Universal arrays may be used to rapidly detect low abundance mutations in any gene of interest.

© 1999 Academic Press

*Corresponding author

Keywords: zip-code addressing; DNA hybridization; thermostable DNA ligase; ligase detection reaction; single nucleotide polymorphism (SNP)

Introduction

Cancers arise from the accumulation of mutations in genes controling the cell cycle, apoptosis, and genome integrity. These mutations may be inherited or somatic, arising from exposure to environmental factors or from malfunctions in DNA replication and repair machinery (Fearon, 1997; Fearon & Vogelstein, 1990; Liu et al., 1996; Perera, 1997). Oncogenes may be activated by point mutations, translocation, or gene amplification, while tumor suppressor genes may be inactivated by point mutations, frameshift mutations

and deletions (Bishop, 1991; Da Costa et al., 1996; Venitt, 1996). A major hurdle to detecting mutations in these genes is that, in primary tumors, normal stromal cell contamination can be as high as 70% of total cells, and thus a mutation present in only one of the two chromosomes of a tumor cell may represent as little as 15% of the DNA sequence present in a sample for that gene. Thus, there is an urgent need to develop technology that can identify accurately one or more low abundance mutations, at multiple adjacent, nearby, and distal loci in a large number of genes.

The advent of DNA arrays has resulted in a paradigm shift in detecting sequence variations and monitoring gene expression levels on a genomic scale (Beattie et al., 1995; Brown & Botstein, 1999; Chee et al., 1996; Cronin et al., 1996; DeRisi et al., 1996; Drobyshev et al., 1997; Eggers et al., 1994; Gunderson et al., 1998; Guo et al., 1994; Hacia, 1999; Hacia et al., 1996; Kozal et al., 1996;

Pease *et al.*, 1994; Schena *et al.*, 1996; Shalon *et al.*, 1996; Southern *et al.*, 1999; Yershov *et al.*, 1996; Zhu *et al.*, 1998). DNA chips designed to distinguish single nucleotide differences are generally based on hybridization of labeled targets (Beattie *et al.*, 1995; Chee *et al.*, 1996; Cronin *et al.*, 1996; Drobyshev *et al.*, 1997; Eggers *et al.*, 1994; Guo *et al.*, 1994; Hacia *et al.*, 1996; Kozal *et al.*, 1996; Parinov *et al.*, 1996; Sapolsky *et al.*, 1999; Wang *et al.*, 1998; Yershov *et al.*, 1996) or polymerase extension of arrayed primers (Lockley *et al.*, 1997; Nikiforov *et al.*, 1994; Pastinen *et al.*, 1997; Shumaker *et al.*, 1996). While DNA chips based on these two formats can confirm a known sequence, the similarities in hybridization profiles create ambiguities in distinguishing heterozygous from homozygous alleles (Beattie *et al.*, 1995; Chee *et al.*, 1996; Eggers *et al.*, 1994; Kozal *et al.*, 1996; Southern, 1996; Wang *et al.*, 1998). To overcome this problem, several methods have been proposed, including the use of: (i) two-color fluorescence analysis (Hacia *et al.*, 1996, 1998a); (ii) a tiling strategy that uses 40 overlapping addresses for each known polymorphism (Cronin *et al.*, 1996); (iii) incorporation of nucleotide analogues in the array sequence (Guo *et al.*, 1997; Hacia *et al.*, 1998b); and (iv) adjacent co-hybridized oligonucleotides (Drobyshev *et al.*, 1997; Gentalen & Chee, 1999, Yershov *et al.*, 1996). A recent side-by-side comparison revealed that the use of hybridization chips for nucleotide discrimination gave an order of magnitude higher background than was observed with the primer extension approach, resulting in an increased likelihood of false positive identifications (Pastinen *et al.*, 1997). Nevertheless, solid-phase primer extension can also generate false positive signals from mononucleotide repeat sequences, template-dependent errors, and template-independent errors (Nikiforov *et al.*, 1994; Shumaker *et al.*, 1996). In addition, neither of these two types of arrays can detect cancer mutations when these are present in a minority of the total target DNA.

Over the past few years, our laboratories have pursued an alternate strategy in DNA array design. In concert with polymerase chain reaction/ligase detection reaction (PCR/LDR) assays carried out in solution (Barany, 1991a,b; Belgrader *et al.*, 1996; Day *et al.*, 1995, 1996; Khanna *et al.*, 1999), our array concept allows for accurate identification of mutations and single nucleotide polymorphisms (SNPs). Primary PCR amplification of the gene of interest is followed by LDR, which uses a thermostable *Tth* DNA ligase that links two adjacent oligonucleotides annealed to a complementary target if and only if the nucleotides are perfectly base-paired at the junction (Figure 1(a)). Since a single-base mismatch prevents ligation, it is possible to distinguish mutations with exquisite specificity, even at low abundance (Khanna *et al.*, 1999). Furthermore, such assays are ideal for multiplexing, since several primer sets can ligate along a gene without the interference encountered in polymerase-based assays (Belgrader *et al.*, 1996; Day *et al.*,

1995; Khanna *et al.*, 1999). High-throughput detection of specific multiplexed LDR products is then achieved *via* divergent sequences termed "zip-code" complements which guide each LDR product to a designated zip-code address on a DNA array (Figure 1(b)). This concept is analogous to molecular tags developed for bacterial and yeast genetics (Hensel *et al.*, 1995; Shoemaker *et al.*, 1996). Based on recent multiplexed PCR/LDR results from our laboratory, the new approach should allow detection of: (i) dozens to hundreds of polymorphisms in a single-tube multiplex format; (ii) small insertions and deletions in repeat sequences; and (iii) low abundance mutations in a background of normal DNA (Khanna *et al.*, 1999, and unpublished results).

Results and Discussion

Zip-code concept and design

Our approach uses microarrays of unique 24-base oligonucleotides that are coupled to a three-dimensional polymer at known locations. These 24-mers or zip-codes (Table 1) hybridize specifically to molecules containing sequences that are complementary to the zip-codes. By linking the zip-code complements to fluorescent primers *via* a tandem PCR/LDR strategy, zip-code microarrays can be used to assess the presence and abundance of mutations in biological specimens. Importantly, because the zip-codes represent unique artificial sequences, zip-code microarrays can be used as a universal platform for molecular recognition simply by changing the gene-specific sequences linked to the zip-code complements.

Each zip-code sequence is composed of six tetramers (designed as described below) such that the full-length 24-mers have similar t_m values. The 256 (4^4) possible combinations in which the four bases can be arranged as tetramers were reduced to a set of 36; these were chosen such that each tetramer differed from all others by at least two bases (Figure 2). Tetramer complements, as well as tetramers that would result in self-pairing or hairpin formation of the zip-codes, were eliminated. Furthermore, tetramers that were palindromic, e.g. TCGA, or repetitive, e.g. CACA, were excluded (diagonally hatched boxes in Figure 2). The indicated set of 36 tetramers represents just one of the possible sets that can be created; alternative sets can be developed by starting in any of the unused light gray boxes (Figure 2).

Six tetramers were chosen from the larger set of 36 for use in designing the zip-codes for the prototype array. These six tetramers were combined such that each zip-code differs from all others by at least three alternating tetramer units (Table 1). This ensures that each zip-code differs from all other zip-codes by at least six bases, thus preventing even the closest zip-code sequences from cross-hybridizing. The t_m values of correct hybridizations range from 70 °C to 82 °C and are at least 24 deg. C
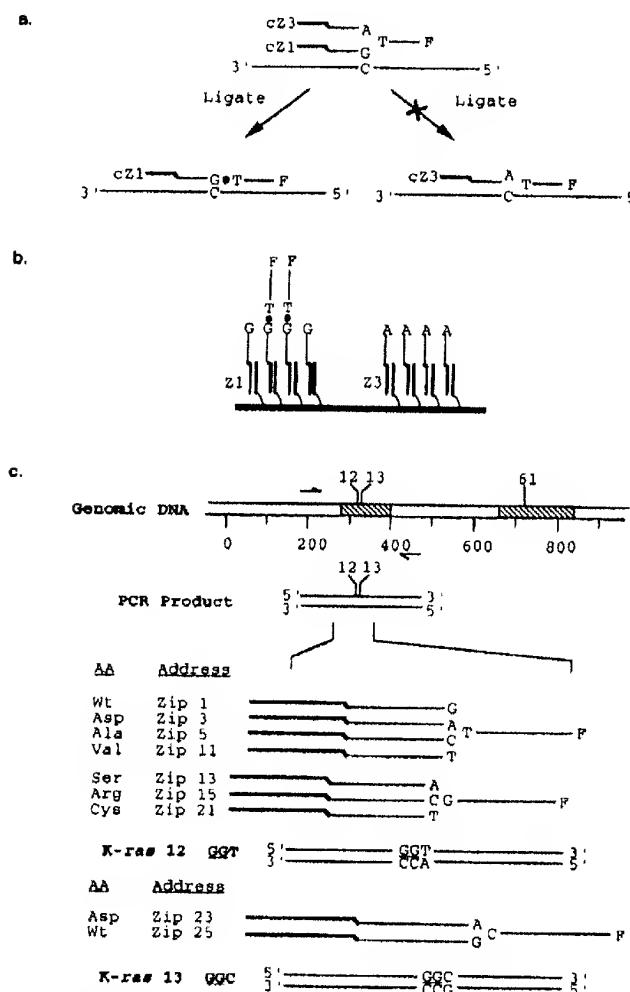
Figure 1. Scheme for PCR/LDR detection of mutations using an addressable array. (a) Schematic representation of LDR primers used to distinguish mutations. Each allele-specific primer contains an addressable sequence complement (cZ1 or cZ3) on the 5'-end and the discriminating base on the 3'-end. The common LDR primer is phosphorylated on the 5'-end and contains a fluorescent label on the 3'-end. The primers hybridize adjacent to each other on target DNA, and the nick will be sealed by the ligase if and only if there is perfect complementarity at the junction. (b) The presence and type of mutation is determined by hybridizing the contents of an LDR to an addressable DNA array The zip-code sequences are designed to be sufficiently different, so that only primers containing the correct complement to a given zip-code will remain bound at that address. (c) Schematic representation of chromosomal DNA containing the K-*ras* gene. Exons are shaded and the positions of codons 12 and 13 are shown. Exon-specific primers were used to selectively amplify K-*ras* DNA flanking codons 12 and 13 Primers were designed for LDR detection of seven possible mutations in these two codons as described in (a)

higher than that of any incorrect hybridization (calculated using Oligo 6 0, Molecular Biology Insights, Inc., Cascade, CO). The concept of using alternating rows and columns of tetramer units may be extended to include all 36 tetramers, hence creating an array with 1296 divergent addresses.

Table 1. Zip-code sequences used in prototype array

Zip#	Tetramer order[a]	Zip-code sequence (5' → 3')[b]
Zip1	1-6-3-2-6-3	TGCG·ACCT·CAGC·ATCG·ACCT·CAGC-spacer-NH$_2$
Zip3	3-6-5-2-2-3	CAGC·ACCT·GACC·ATCG·ATCG·CAGC-spacer-NH$_2$
Zip5	5-6-1-2-4-3	GACC·ACCT·TGCG·ATCG·GGTA·CAGC-spacer-NH$_2$
Zip11	1-4-3-6-6-1	TGCG·GGTA·CAGC·ACCT·ACCT·TGCG-spacer-NH$_2$
Zip13	3-4-5-6-2-1	CAGC·GGTA·GACC·ACCT·ATCG·TGCG-spacer-NH$_2$
Zip15	5-4-1-6-4-1	GACC·GGTA·TGCG·ACCT·GGTA·TGCG-spacer-NH$_2$
Zip21	1-2-3-4-6-5	TGCG·ATCG·CAGC·GGTA·ACCT·GACC-spacer-NH$_2$
Zip23	3-2-5-4-2-5	CAGC·ATCG·GACC·GGTA·ATCG·GACC-spacer-NH$_2$
Zip25	5-2-1-4-4-5	GACC·ATCG·TGCG·GGTA·GGTA·GACC-spacer-NH$_2$

[a] Order of tetramer oligonucleotide segments in the corresponding zip-code sequence. Six tetramers were chosen from the full set of 36 to prepare the zip-codes for the prototype array. The six tetramers which were renumbered for ease of use are: 1, TGCG, 2, ATCG, 3, CAGC; 4, GGTA, 5, GACC, and 6, ACCT. Closely related sequences, (Zip1, 3, 5), (Zip11, 13, 15) and (Zip21, 23, 25) differ at the first, third, and fifth tetramer positions, but are identical at the second, fourth, and sixth tetramer positions.
[b] spacer-NH$_2$ = -O(PO$_2$)O-(CH$_2$CH$_2$O)$_6$-PO$_2$-O(CH$_2$)$_3$NH$_2$

Array preparation

Numerous types of two and three-dimensional matrices were examined with respect to: (i) ease of preparation of the surface; (ii) oligonucleotide loading capacity; (iii) stability to conditions required for coupling of oligonucleotides, as well as for hybridization and washing; and (iv) compatibility with fluorescence detection. Our currently favored methodology to construct zip-code arrays involves initial creation of a lightly crosslinked film of acrylamide/acrylic acid copolymer on a glass solid support; subsequently, the free carboxyl groups dispersed randomly throughout the polymeric surface are activated with N-hydroxysuccinimide, and amine terminated zip-code oligonucleotide probes are added to form covalent amide linkages (Figure 3(a)). The described coupling chemistry is rapid, straightforward, efficient, and amenable to both manual and robotic spotting. Both the activated surfaces and the surfaces with attached oligonucleotides are stable to long-term storage.

Optimization of hybridization conditions

Hybridizations of a fluorescently labeled 70-mer probe onto model zip-code arrays were measured as a function of buffer, metal cofactors, volume, pH, time, and the mechanics of mixing (Table 2). Even with closely related zip-codes, cross-hybridization was negligible or non-existent, with a signal-to-noise ratio of at least 50:1. Our experiments suggest that different zip-codes hybridize at approximately the same rate, i.e. the level of fluorescent signal is relatively uniform when normalized for the amount of oligonucleotide coupled per address (data not shown). Magnesium ion was obligatory to achieve hybridization, and less than 1 fmol of probe could be detected in the presence of this divalent cation (Table 2 and Figure 4). The hybridization signal was doubled upon lowering the pH from 8.0 to 6.0, most likely due to masking of negative charges (hence reducing repulsive interactions with oligonucleotides) arising from uncoupled acrylic acid groups in the bulk polymer

Table 2. Effect of hybridization conditions on hybridization signal

Hybridization buffer	Vol. (µl)	Mixing[a]	Time (minutes)	Relative signal
Buffer A	55	Inter.	30	1
Buffer A minus MgCl$_2$	55	Inter.	30	<0.01
Buffer A	20	Inter.	30	2.5
Buffer B	55	Inter.	30	2
Buffer B	20	Inter.	30	3
Buffer B	55	Cont.	30	4
Buffer B	55	Cont.	60	8
Buffer A + Capped Surface	55	Cont.	60	8
Buffer B minus MgCl$_2$	55	Cont.	60	<0.01
Buffer B	55	Cont.	180	10

Following general procedures described in Materials and Methods, hybridizations were carried out with 1 pmol of FAMcZip13-Prd and 3 × 3 manually spotted arrays. Buffers were: buffer A, 300 mM bicine (pH 8.0), 10 mM MgCl$_2$, 0.1 % SDS; buffer B, 300 mM Mes (pH 6.0), 10 mM MgCl$_2$, 0.1 % SDS.
[a] Mixing was as follows: intermittent (Inter.), manual mixing of the sample once every ten minutes; continuous (Cont.), mixing of sample at 20 rpm in a hybridization oven.
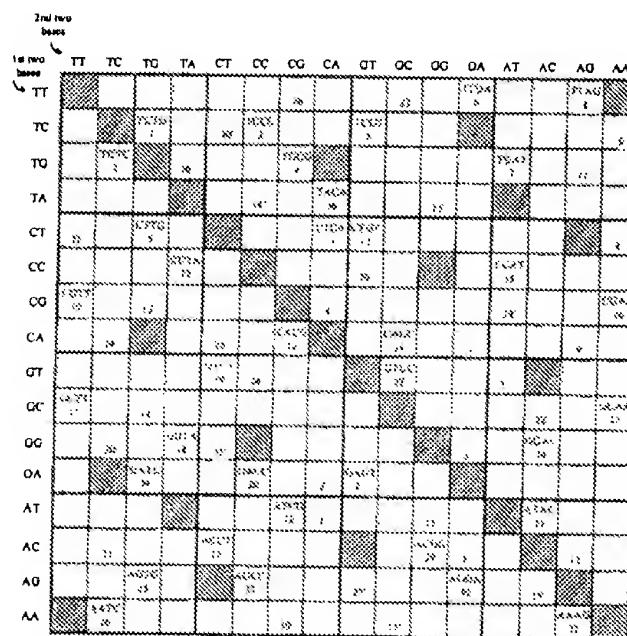
Figure 2. Design of tetramers for use in zip-code arrays. The checkerboard pattern shows all 256 possible tetramers. A given square represents the two bases on the left followed by the two bases on the top of the checkerboard. To be included, each tetramer must differ from all others by at least two bases, and be non-complementary. The chosen tetramers are shown in the white boxes, while their complements are listed as (number)'. Thus, as an example, the complementary sequences GACC (20) and GGTC (20') are mutually exclusive in this scheme. In addition, tetramers that are palindromic, e.g. TCGA (off-diagonal hatched boxes) or repetitive, e.g. CACA (hatched boxes on diagonal from upper left to lower right) have been eliminated. All other sequences which differ from the 36 tetramers by only one base are shaded in light gray. Four potential tetramers were not chosen as they are either all A T or G C bases (open boxes)

matrix. To confirm this hypothesis, the free carboxyl groups on arrays to which zip-codes had already been attached were capped with ethanolamine under standard coupling conditions. Hybridizations of the capped arrays at pH 8.0 gave results comparable to hybridizations at pH 6.0 of the same arrays without capping. Continuous mixing proved to be crucial for obtaining good hybridization, and studies of the time-course led us to choose one hour at 65 °C as standard. Reducing the hybridization volume improved the hybridization signal due to the relative increase in probe concentration. Further improvements may be achieved using specialized small volume hybridization chambers that allow for continuous mixing.

Array hybridization of K-ras LDR products

PCR/LDR amplification coupled with zip-code detection on an addressable array was tested with the K-ras gene as a model system. Exon-specific PCR primers were used to selectively amplify K-ras DNA flanking codons 12 and 13. LDR primers were designed to detect the seven most common mutations found in the K-ras gene in colorectal cancer (Figure 1(c) and Table 3). For example, the second position in codon 12, GGT, coding for glycine, may mutate to GAT, coding for aspartate, which is detected by ligation of the allele-specific primer (containing a zip-code complement, cZip3, on its 5'-end, and a discriminating base, A, on its 3'-end) to a fluorescently labeled common primer (Figure 1(c)).

PCR/LDR was carried out on nine individual DNA samples derived from cell lines or paraffin-embedded tumors containing known K-ras mutations (as described in Materials and Methods). An aliquot (2 µl) was taken from each reaction and electrophoresed on a sequencing apparatus to confirm that LDR was successful (data not shown). Next, the different mutations were distinguished by hybridizing the LDR product mixtures on 3 × 3 addressable DNA arrays (each zip-code address was spotted in quadruplicate), and detecting the
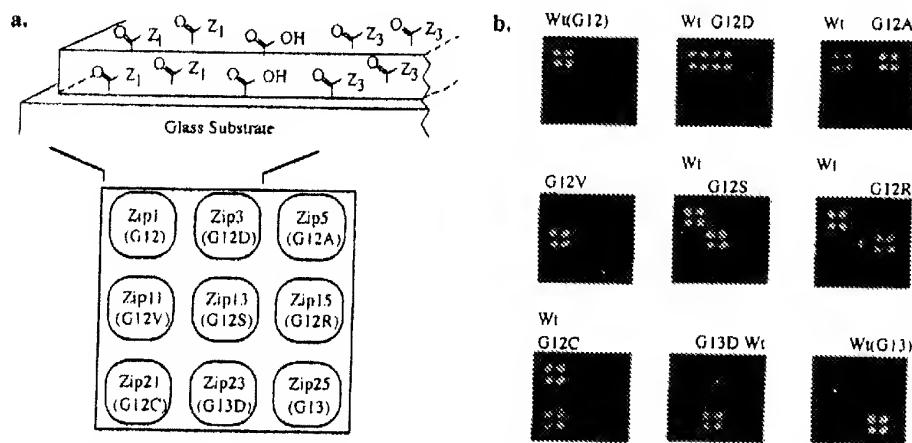
Figure 3. Detection of K-*ras* mutations on a DNA array. (a) Schematic representation of gel-based zip-code array. Glass microscope slides treated with γ-methacryloyloxypropyltrimethoxysilane are used as the substrate for the covalent attachment of an acrylamide/acrylic acid copolymer matrix. Amine-modified zip-code oligonucleotides are coupled to N-hydroxysuccinimide-activated surfaces at discrete locations (see Materials and Methods) Each position in the 3 × 3 grid identifies an individual zip-code address (and corresponding K-*ras* mutation or wild-type sequence). (b) Each robotically spotted array was hybridized with an individual LDR and fluorescent signal detected as described in Materials and Methods using a two second exposure time. All nine arrays identified the correct mutant and/or wild-type for each tumor (G12S, G12R, and G12C) or cell line sample (Wt, G12D, G12A, G12V, and G13D) The small spots seen in some of the panels, e.g. near the center of the panel containing the G13D mutant, are not incorrect hybridizations, but noise due to imperfections in the polymer

positions of fluorescent spots (Figure 3(b)). The wild-type samples, Wt(G12) and Wt(G13), each displayed four equal hybridization signals at Zip1 and Zip25, respectively, as expected. The mutant samples each displayed hybridization signals corresponding to the mutant, as well as for the wild-type DNA present in the cell line or tumor. The sole exception to this was the G12V sample, which

was derived from a cell line (SW620) homozygous for the G12V K-*ras* allele. The experiment was repeated several times, using both manually and robotically spotted arrays, and LDR primers labeled with either fluorescein or Texas Red. False-positive or false-negative signals were not encountered in any of these experiments. A minimal amount of noise seen on the arrays can be attribu-

Fmol of Fluorescein-Labeled 70-mer Complement to Zip13
Mixed with 9000 fmol of LDR Primers
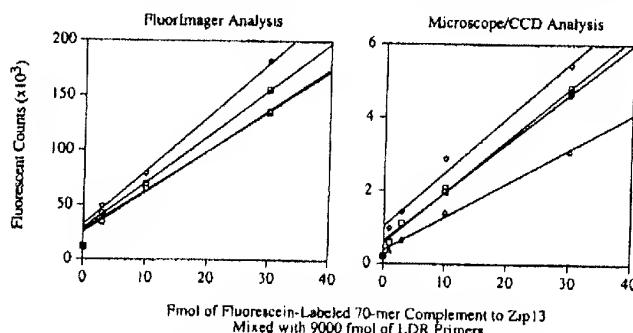
Figure 4. Determination of zip-code array capture sensitivity using two different detection instruments. Quadruplicate hybridizations were carried out on manually spotted arrays as described in Materials and Methods. The graphs depict quantification of the amount of captured 70-mer complement using either a fluorimager (left) or an epifluorescence microscope/CCD (right) Each symbol represents hybridizations to an individual array. The filled square on each graph is the average of the backgrounds from all four arrays.

Table 3. Primers designed for K-ras mutation detection by PCR/LDR/array hybridization

Primer	Sequence (5′ → 3′)
K-ras exon 1 forward	ATAAGGCCTGCTGAAAATGACTGAA
K-ras exon 1 reverse	CTGCACCAGTAATATGCATATTAAAACAAG
cZip1-K-ras c12 2WtG	GCTGAGGTCGATGCTGAGGTCGCAAAACTTGTGGTAGTTGGAGCTGG
cZip3-K-ras c12 2D	GCTGCGATCGATGGTCAGGTGCTGAAACTTGTGGTAGTTGGAGCTGA
cZip5-K-ras c12.2A	GCTGTACCCGATCGCAAGGTGGTCAAACTTGTGGTAGTTGGAGCTGC
cZip11-K-ras c12.2V	CGCAAGGTAGGTGCTGTACCCGCAAAACTTGTGGTAGTTGGAGCTGT
K-ras c12 Com-2	pTGGCGTAGGCAAGAGTGCC1-fluorescein
	pTGGCGTAGGCAAGAGTGCCT-Texas Red
cZip13-K-ras c12 1S	CGCACGATAGGTGGTCTACCGCTGATATAAACTTGTGGTAGTTGGAGCTA
cZip15-K-ras c12 1R	CGCATACCAGGTCGCATACCGGTCATATAAACTTGTGGTAGTTGGAGCTC
cZip21-K-ras c12 1C	GGTCAGGTTACCGCTGCGATCGCAATATAAACTTGTGGTAGTTGGAGCTT
K-ras c12 Com-1	pGTGGCGTAGGCAAGAGTGCC-fluorescein
	pGTGGCGTAGGCAAGAGTGCC-Texas Red
cZip23-K-ras c13 4D	GGTCCGATTACCGGTCCGATGCTGTGTGGTAGTTGGAGCTGGTGA
cZip25-K-ras c13 4WtG	GGTCTACCTACCCGCACGATGGTCTGTGGTAGTTGGAGCTGGTGG
K-ras c13 Com-4	pCGTAGGCAAGAGTGCCTTGAC-fluorescein
	pCGTAGGCAAGAGTGCCTTGAC-Texas Red

The PCR primers were specifically designed to amplify exon 1 of K-ras without co-amplifying N and H-ras. The allele-specific LDR primers contained 24-mer zip-code complement sequences on their 5′-ends (bold) and the discriminating bases on their 3′-ends (underlined). The common LDR primers contained 5′-phosphate groups and either a fluorescein or a Texas Red label on their 3′-ends.

ted to dust, scratches, and/or small bubbles in the polymer. These flaws are readily recognized because they are weak and sporadic, rather than reproducing the quadruplicate spotting pattern; we expect such noise will be minimized with more stringent manufacturing conditions. Ultimately, these protocols are amenable to quantifying the relative amounts of each allele, and work is currently in progress to convert our quantitative PCR/LDR protocols for K-ras mutations from gel-based detection to array-based detection (unpublished results).

Array capture sensitivity

After an LDR, the successfully ligated and fluorescently labeled LDR product competes with an excess of unligated discriminating primer for hybridization to the correct zip-code address on the array. To determine capture sensitivity, DNA arrays were hybridized in quadruplicate, under standard conditions, with from 100 amol (= 1/90,000) to 30 (= 1/300) fmol of a labeled synthetic 70-mer, FAMcZip13-Prd (this simulates a full-length LDR product; see Materials and Methods for the sequence), in the presence of a full set of K-ras LDR primers (combined total of 9000 fmol of discriminating and common primers). Array analyses with a FluorImager (Figure 4, left-side) indicate that a signal-to-noise ratio of greater than 3:1 can be achieved when starting with a minimum of 3 fmol (= 1/3,000) of FAMcZip13-Prd-labeled probe in the presence of 4500 fmol of FAM-labeled LDR primers and 4500 fmol of zip-code complement primers in the hybridization solution. Results using microscope/CCD instrumentation to quantify fluorescence were even more striking: a 3:1 signal-to-noise ratio was maintained starting with 1 fmol (= 1/9,000) of labeled product (Figure 4, right-hand

side) on three out of the four arrays; the signal to noise was 2:1 on the fourth array. For a given array, with fluorescence quantified by either instrument, the captured counts varied linearly with the amount of labeled FAMcZip13-Prd added. Rehybridization of the same probe, at the same concentration, to the same array, was reproducible within ±5% (data not shown). Variations in fluorescent signal between arrays may reflect variations in the amount of zip-code oligonucleotide coupled, due to the inherent inaccuracies of manual spotting and/or variations in polymer uniformity.

Detection of low abundance mutations by PCR/LDR/array hybridization

To determine the limit of detection of low-level mutations in wild-type DNA using PCR/LDR/array hybridization, a dilution series was set up and analyzed. PCR-amplified pure G12V DNA was diluted into wild-type K-ras DNA in ratios ranging from 1:20 to 1:500. Duplicate LDRs were carried out on 2000 fmol of total DNA, using a two-primer set consisting of 2000 fmol each of the discriminating and common primers for the G12V mutation. It proved possible to quantify a positive hybridization signal at a dilution of 1:200 with a signal-to-noise ratio of 2:1 (Figure 5). A signal was distinguishable even at a dilution of 1:500, although noise levels due to dust or bubbles in the polymer prevented us from accurately quantifying the results. A control of pure wild-type DNA showed no hybridization signal. These results indicate clearly that zip-code array hybridization, when coupled with PCR/LDR, may detect polymorphisms present at less than 1% of the total DNA. These results are consistent with our earlier work showing that PCR/LDR, using a 26-primer set and analyses based on gel electrophoreses of
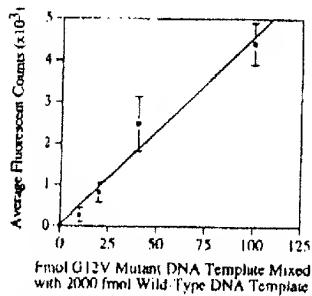
Figure 5. Detection of minority K-*ras* mutant DNA in a majority of wild-type DNA using PCR/LDR with zip-code array capture. DNA from cell line SW620, containing the G12V mutation, and DNA from normal lymphocytes were PCR amplified in exon 1 of the K-*ras* gene. Mixtures containing 10, 20, 40, or 100 fmol of G12V-amplified fragment plus 2000 fmol of PCR-amplified wild-type fragment were prepared, and the presence of mutant DNA determined by LDR using primers specific for the G12V mutation (2000 fmol each of discriminating and common primer). Images were collected by CCD using exposure times from five to 25 seconds. Data were normalized by dividing fluorescent signal intensity by acquisition time. Each data point represents the average hybridization signal from four independent robotically spotted arrays. The average background signal from all four spots at each address following hybridization of pure wild-type control (880 average fluorescent counts) was subtracted from the mutant signal.

products, can detect any K-*ras* mutation in the presence of up to a 500-fold excess of wild-type, with a signal-to-noise ratio of at least 3:1 (Khanna *et al.*, 1999).

Comparison of universal array to gene-specific arrays

Our approach to mutation detection has three orthogonal components: (i) primary PCR amplification; (ii) solution-phase LDR detection; and (iii) solid-phase hybridization capture. Therefore, background signal from each step can be minimized and, consequently, the overall sensitivity and accuracy of our method are significantly enhanced over those provided by other strategies. For example, hybridization of labeled target methods require: (i) multiple rounds of PCR or PCR/T7 transcription; (ii) processing of PCR amplified products to fragment them or render them single-stranded; and (iii) lengthy hybridization periods (ten hours or more) which limits throughput (Chee *et al.*, 1996; Cronin *et al.*, 1996; Guo *et al.*, 1994; Hacia *et al.*, 1996; Schena *et al.*, 1996; Shalon *et al.*, 1996; Wang *et al.*, 1998). Additionally, since the immobilized probes on the aforementioned arrays have a wide range of t_m values, it is necessary to perform the

hybridizations at temperatures from 0 °C to 44 °C. The result is increased background noise and false signals due to mismatch hybridization and non-specific binding, for example, on small insertions and deletions in repeat sequences (Cronin *et al.*, 1996; Hacia *et al.*, 1996; Southern, 1996; Wang *et al.*, 1998). In contrast, our approach allows multiplexed PCR in a single reaction (Belgrader *et al.*, 1996), does not require an additional step to convert product into single-stranded form, and can readily distinguish all point mutations including slippage in repeat sequences (Day *et al.*, 1995; Khanna *et al.*, 1999). Alternative DNA arrays suffer from differential hybridization efficiencies due to either sequence variation or to the amount of target present in the sample. By using our approach of designing divergent zip-code sequences with similar thermodynamic properties, hybridizations can be carried out at 65 °C, resulting in a more stringent and rapid hybridization. The decoupling of the hybridization step from the mutation detection stage offers the prospect of quantification of LDR products, as we have already achieved using gel-based LDR detection (Khanna *et al.*, 1999).

Arrays spotted on polymer surfaces provide substantial improvements in signal capture, as compared with arrays spotted or synthesized *in situ* directly on glass surfaces (Drobyshev *et al.*, 1997; Parinov *et al.*, 1996; Yershov *et al.*, 1996). However, the polymers described by others are limited to using 8 to 10-mer addresses, while our polymeric surface readily allows 24-mer zip-codes to penetrate and couple covalently. Moreover, LDR products of length 60 to 75 nucleotide bases are also found to penetrate and subsequently hybridize to the correct address. As additional advantages, our polymer gives little or no background fluorescence and does not exhibit non-specific binding of fluorescently labeled oligonucleotides. Finally, zip-codes spotted and coupled covalently at a discrete address do not "bleed over" to neighboring spots, hence obviating the need to physically segregate sites, e.g. by cutting gel pads.

Summary and Conclusions

Here, we describe a strategy for high-throughput mutation detection which differs substantially from other array-based detection systems presented previously in the literature. In concert with a polymerase chain reaction/ligase detection reaction (PCR/LDR) assay carried out in solution, our array allows for accurate detection of single base mutations, whether inherited and present as 50% of the sequence for that gene, or sporadic and present at 1% or less of the wild-type sequence. We achieve this sensitivity because thermostable DNA ligase provides the specificity of mutation discrimination, while the divergent addressable portions (zip-codes) of our LDR primers guide each LDR product to a designated address on the DNA array. Since the zip-code sequences remain con-

stant and their complements can be appended to any set of LDR primers, our zip-code arrays are universal. Thus, a single array design can be programmed to detect a wide range of genetic mutations.

Robust methods for the rapid detection of mutations at numerous potential sites in multiple genes hold great promise to improve the diagnosis and treatment of cancer patients. Non-invasive tests for mutational analysis of shed cells in saliva, sputum, urine, and stool could significantly simplify and improve the surveillance of high risk populations, reduce the cost and discomfort of endoscopic testing, thus leading to more effective diagnosis of cancer in its early, curable stage. Although the feasibility of detecting shed mutations has been demonstrated clearly in patients with known and genetically characterized tumors (Caldas *et al.*, 1994; Hasegawa *et al.*, 1995; Nollau *et al.*, 1996; Sidransky *et al.*, 1992; Wu *et al.*, 1994), effective presymptomatic screening will require that a myriad of potential low frequency mutations be identified with minimal false-positive and false-negative signals. Furthermore, the integration of technologies for determining genetic changes within a tumor with clinical information about the likelihood of response to therapy could radically alter how patients with more advanced tumors are selected for treatment. Identification and validation of reliable genetic markers will require that many candidate genes be tested in large-scale clinical trials. While costly microfabricated chips can be manufactured with over 100,000 addresses, none of them has as yet demonstrated a capability to detect low abundance mutations (Chee *et al.*, 1996; Hacia *et al.*, 1996; Kozal *et al.*, 1996; Sapolsky *et al.*, 1999; Wang *et al.*, 1998), as required to accurately score mutation profiles in such clinical trials. The universal zip-code array approach introduced here has the potential to allow rapid and reliable identification of low abundance mutations in multiple codons in numerous genes. As new therapies targeted to specific genes or specific mutant proteins are developed, the importance of rapid and accurate high-throughput genetic testing will undoubtedly increase.

Materials and Methods

Oligonucleotide synthesis and purification

Oligonucleotides were obtained as custom synthesis products from IDT, Inc (Coralville, IA), or synthesized in-house on an ABI 394 DNA Synthesizer (PE Biosystems Inc., Foster City, CA) using standard phosphoramidite chemistry. Spacer phosphoramidite 18, 3'-amino-modifier C3 CPG, and 3'-fluorescein CPG were purchased from Glen Research (Sterling, VA). All other reagents were purchased from PE Biosystems. Oligonucleotides with 3'-amino modifications and/or fluorescent labels were cleaved from the supports by treatment with concentrated aqueous NH_4OH for two hours at 25°C, and deprotection continued in solution for 24 hours at 25°C. Texas Red labeling was achieved by adding 150 μl of

0.2 M $NaHCO_3$ and 200 μg of oligonucleotide to tubes containing a solution of 500 μg of Texas Red-X succinimidyl ester (Molecular Probes; Eugene, OR) in 28 μl of anhydrous DMF. Following overnight stirring at 25°C, the majority of unreacted label was removed by the addition of 20 μl of 3 M NaCl and 500 μl of cold ethanol, chilling in a dry ice/ethanol bath for 30 minutes, and centrifuging at 12,000 *g* for 30 minutes. The supernatants were removed, the pelleted oligonucleotides were washed with 100 μl of 70% ethanol, and dried. FAMcZip13-Prd, a fluorescein-labeled 70-mer that simulates a full-length LDR product containing the complementary sequence to Zip13, was synthesized on 1000 Å pore-size CPG. The sequence was: 5'-fluorescein-CGCACGATAGG **TGGTCTACCGCTG-ATATAAACTTGTGGTAGTTGG-AGCTAGTGGCGTAGGCAAGAGTGCC**-3' (the zip-code complement is in bold).

Both labeled and unlabeled oligonucleotides were purified by electrophoresis on denaturing 12% polyacrylamide gels. Bands were visualized by UV shadowing, excised from the gel, and eluted overnight in 0.5 M NaCl, 5 mM EDTA (pH 8.0) at 37°C. Oligonucleotide solutions were desalted on C18 Sep-Paks (Waters Corporation; Milford, MA) according to the manufacturer's instructions, following which the oligonucleotides were concentrated to dryness (Speed-Vac) and stored at −20°C.

DNA extraction from cell lines

Cell lines of known K-*ras* genotype (HT29, wild-type; SW1116, G12A; LS180, G12D; SW620, G12V; DLD1, G13D) were grown in RPMI culture media with 10% fetal bovine serum. Harvested cells (~10^7) were resuspended in DNA extraction buffer (10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 mM EDTA (pH 8.0), 0.5% (w/v) SDS, 200 μg/ml proteinase K) and incubated at 37°C for four hours. A 30% volume of 6 M NaCl was added and the mixture was centrifuged. The supernatant was transferred to a clean tube and the DNA was pelleted through the addition of three volumes of ethanol, chilling on dry ice, and centrifugation. The pellet was washed with 70% ethanol and resuspended in 10 mM Tris-HCl (pH 7.5), 2 mM EDTA (pH 8.0).

DNA extraction from paraffin sections

Tissue sections (10 μm) were cut from paraffin-embedded colon tumors. Samples were deparaffinized *via* sequential extraction with xylene, ethanol, and acetone, and dried under vacuum. The DNA in the pellets was purified using a QIAamp Tissue Kit (Qiagen, Chatsworth, CA).

Polymer coated slides

Microscope slides (Fisher Scientific, precleaned, 3 in × 1 in × 1.2 mm) were immersed in 2% γ-methacryloyloxypropyltrimethoxysilane, 0.2% triethylamine in $CHCl_3$ for 30 minutes at 25°C, and then washed with $CHCl_3$ (two washes of 15 minutes). A monomer solution (20 μl of 8% acrylamide, 2% acrylic acid, 0.02% *N,N*'-methylene-bisacrylamide (500:1 ratio of monomers:crosslinker), 0.8% ammonium persulfate radical polymerization initiator) was deposited on one end of the slides and spread out with the aid of a cover-slip (24 mm × 50 mm) that had been previously silanized (5% $(CH_3)_2SiCl_2$ in $CHCl_3$). Polymerization was achieved

by heating the slides on a 70 °C hotplate for 4.5 minutes Upon removal of the slides from the hotplate, the coverslips were immediately peeled off with the aid of a single-edge razor blade. The coated slides were rinsed with deionized water, allowed to dry in open atmosphere, and stored under ambient conditions

Zip-code arrays

Polymer-coated slides were pre-activated by immersing them for 30 minutes at 25 °C in a solution of 0 1 M 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride plus 20 mM N-hydroxysuccinimide in 0.1 M K_2HPO_4/KH_2PO_4 (pH 6.0). The activated slides were rinsed with water, and then dried in a 65 °C oven; they were stable upon storage for six months or longer at 25 °C in a desiccator over Drierite.

For manual spotting, 0 2 μl aliquots were taken with a Rainin Pipetman from stock solutions (500 μM) of zipcode oligonucleotides in 0.2 M K_2HPO_4/KH_2PO_4 (pH 8 3), and deposited in a 3 × 3 array onto the pre-activated polymeric surfaces. The resulting arrays were incubated for one hour at 65 °C in humidified chambers containing water/formamide (1:1). For robotic spotting, 10-50 nL aliquots of zip-code oligonucleotides (1.5 mM in the same buffer) were deposited at 25 °C on the pre-activated surfaces by using a robot (PE Biosystems, "inhouse" design) equipped with a quill-type spotter in a controled atmosphere chamber. Two pairs of 3 × 3 arrays were spotted on each slide, with addresses consisting of groups of four spots. Following spotting using either method, uncoupled oligonucleotides were removed from the polymer surfaces by soaking the slides in 300 mM bicine (pH 8 0), 300 mM NaCl, 0.1 % SDS, for 30 minutes at 65 °C, rinsing with water, and drying The arrays were stored at 25 °C in slide boxes until needed.

PCR amplification of K-*ras* DNA samples

PCR amplifications were carried out under paraffin oil in 20 μl reaction mixtures containing 10 mM Tris-HCl (pH 8 3), 1 5 mM $MgCl_2$, 50 mM KCl, 800 μM dNTPs, 2 5 μM forward and reverse primers (12 5 pmol of each primer; Table 3), and 1-50 ng of genomic DNA extracted from paraffin-embedded tumors or from cell lines. Following a two minute denaturation step at 94 °C, 0.2 unit of Taq DNA polymerase (PE Biosystems) was added Amplification was achieved by thermally cycling for 40 rounds of 94 °C for 15 seconds and 60 °C for two minutes, followed by a final elongation at 65 °C for five minutes Following PCR, 1 μl of proteinase K (18 mg/ml) was added, and reactions were heated to 70 °C for ten minutes and then quenched at 95 °C for 15 minutes One microliter of each PCR product was analyzed on a 3 % agarose gel to verify the presence of amplification product of the expected size.

LDR of K-*ras* DNA samples

LDR was carried out under paraffin oil in 20 μl volumes containing 20 mM Tris-HCl (pH 8 5), 5 mM $MgCl_2$, 100 mM KCl, 10 mM DTT, 1 mM NAD^+, 8 pmol of total LDR primers (500 fmol each of discriminating primers + 4 pmol of fluorescently labeled common primers), and 1 pmol of PCR products from cell line or tumor samples Two primer mixes were prepared, each containing the seven mutation-specific primers, the three common primers, and either the wild-type discriminat-

ing primer for codon 12 or that for codon 13 (Figure 1(c) and Table 3).

The reaction mixtures were pre-heated for two minutes at 94 °C, and then 25 fmol of wild-type Tth DNA ligase was added The LDRs were cycled for 20 rounds of 94 °C for 30 seconds and 65 °C for four minutes An aliquot of 2 μl of each reaction was mixed with 2 μl of gel loading buffer (8 % blue dextran, 50 mM EDTA (pH 8.0), formamide (1:5)), denatured at 94 °C for two minutes, and chilled on ice; 1 μl of each mixture was loaded onto a denaturing 10 % polyacrylamide gel and electrophoresed on an ABI 377 DNA sequencer at 1500 volts.

Hybridization of K-*ras* LDR products to DNA arrays

The LDRs (17 μl) were diluted with 40 μl of 1.4× hybridization buffer to produce a final buffer concentration of 300 mM Mes (pH 6.0), 10 mM $MgCl_2$, 0.1 % SDS, denatured at 94 °C for three minutes, and chilled on ice. Arrays were pre-incubated for 15 minutes at 25 °C in 1× hybridization buffer. Coverwells (Grace, Inc; Sunriver, OR) were filled with the diluted LDRs and attached to the arrays. The arrays were placed in humidified culture tubes and incubated for one hour at 65 °C and 20 rpm in a rotating hybridization oven. Following hybridization, the arrays were washed in 300 mM bicine (pH 8 0), 10 mM $MgCl_2$, 0.1 % SDS for ten minutes at 25 °C. Fluorescent signals were measured using a microscope/CCD (see below).

Hybridization of synthetic LDR products to DNA arrays

Quadruplicate hybridization mixtures were prepared containing 100 amol, 1fmol, 3 fmol, 10 fmol, or 30 fmol of FAMcZip13-Prd (a synthetic 70-mer LDR product complementary to zip-code 13) combined with 4500 fmol of total fluorescein-labeled common LDR primers and 9 × 500 fmol of each unlabeled, zip-code-containing discriminating LDR primer in 55 μl of 300 mM Mes (pH 6 0), 10 mM $MgCl_2$, 0.1 % SDS. Hybridizations were conducted according to the protocol described above, and FluorImager as well as epifluorescence microscopy data were acquired and analyzed (see below)

LDR and hybridization of G12V/G12 dilution series to DNA arrays

These experiments were carried out in a volume of 20 μl. The PCR-amplified SW620 cell line DNA containing the G12V mutation was diluted from 5 nM (100 fmol = 1/20) to 0.050 nM (1 fmol = 1/2000) in LDR mixtures containing 100 nM (2000 fmol) of wild-type (G12) DNA and 100 nM (2000 fmol) of both G12V-discriminating primer and Texas Red-labeled common primer The LDR and hybridization proceeded as above, and imaging on the microscope/CCD was carried out as detailed below

Image analysis

Arrays were imaged using a Molecular Dynamics FluorImager 595 (Sunnyvale, CA) or an Olympus AX70 epifluorescence microscope (Melville, NY) equipped with a Princeton Instruments TE/CCD-512 TKBM1 camera (Trenton, NJ). For analysis of fluorescein-labeled probes on the FluorImager, the 488 nm excitation was used with

a 530/30 emission filter. The spatial resolution of scans was 100 μm per pixel. The resulting images were analyzed using ImageQuaNT software provided with the instrument. The epifluorescence microscope was equipped with a 100 W mercury lamp, a FITC filter cube (excitation 480/40, dichroic beam splitter 505, emission 535/50), a Texas Red filter cube (excitation 560/55, dichroic beam splitter 595, emission 645/75), and a 100 mm macro objective. The macro objective allows illumination of an object field up to 15 mm in diameter and projects a 7 mm × 7 mm area of the array onto the 12.3 mm × 12.3 mm matrix of the CCD. Images were collected in 16-bit mode using the Winview32 software provided with the camera. Analysis was performed using Scion Image (Scion Corporation, Frederick, MD).

Acknowledgements

References

Barany, F. (1991a). Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc Natl Acad Sci. USA*, 88, 189-193.

Barany, F. (1991b). The ligase chain reaction (LCR) in a PCR world. *PCR Methods Appl* 1, 5-16.

Beattie, K. L., Beattie, W. G., Meng, L., Turner, S. L., Coral-Vazquez, R., Smith, D. D., McIntyre, P. M. & Dao, D. D (1995). Advances in genosensor research. *Clin Chem* 41, 700-706.

Belgrader, P., Marino, M., Lubin, M. & Barany, F. (1996). A multiplex PCR-ligase detection reaction assay for human identity testing. *Genome Sci. Technol.* 1, 77-87

Bishop, J. M. (1991) Molecular themes in oncogenesis. *Cell*, 64, 235-248.

Brown, P. O. & Botstein, D (1999) Exploring the new world of the genome with DNA microarrays. *Nature Genet* 21, 33-37.

Caldas, C., Hahn, S. A., Hruban, R. H., Redston, M. S., Yeo, C. J. & Kern, S. E. (1994). Detection of K-*ras* mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. *Cancer Res* 54, 3568-3573

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S. & Fodor, S. P. (1996). Accessing genetic information with high-density DNA arrays. *Science*, 274, 610-614.

Cronin, M. T., Fucini, R. V., Kim, S. M., Masino, R. S., Wespi, R. M. & Miyada, C. G. (1996) Cystic fibrosis

mutation detection by hybridization to light-generated DNA probe arrays. *Human Mutat.* 7, 244-255.

Da Costa, L. T., Jen, J., He, T. C., Chan, T. A., Kinzler, K. W. & Vogelstein, B. (1996). Converting cancer genes into killer genes. *Proc Natl Acad Sci. USA*, 93, 4192-4196.

Day, D. J., Speiser, P. W., White, P. C. & Barany, F. (1995). Detection of steroid 21 hydroxylase alleles using gene-specific PCR and a multiplexed ligation detection reaction. *Genomics*, 29, 152-162.

Day, D. J., Speiser, P. W., Schulze, E., Bettendorf, M., Fitness, J., Barany, F. & White, P. C. (1996). Identification of non-amplifying CYP21 genes when using PCR-based diagnosis of 21-hydroxylase deficiency in congenital adrenal hyperplasia (CAH) affected pedigrees. *Hum. Mol. Genet* 5, 2039-2048.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A. & Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer *Nature Genet.* 14, 457-460.

Drobyshev, A., Mologina, N., Shik, V., Pobedimskaya, D., Yershov, G. & Mirzabekov, A. (1997). Sequence analysis by hybridization with oligonucleotide microchip: identification of β-thalassemia mutations *Gene*, 188, 45-52.

Eggers, M., Hogan, M., Reich, R. K., Lamture, J., Ehrlich, D., Hollis, M., Kosicki, B., Powdrill, T., Beattie, K., Smith, S., Varma, R., Gangadharam, R., Mallik, A., Burke, B. & Wallace, D (1994). A microchip for quantitative detection of molecules utilizing luminescent and radioisotope reporter groups. *BioTechniques*, 17, 516-525.

Fearon, E. R. (1997). Human cancer syndromes: clues to the origin and nature of cancer. *Science*, 278, 1043-1050.

Fearon, E. R. & Vogelstein, B (1990). A genetic model for colorectal tumorigenesis. *Cell*, 61, 759-767

Gentalen, E. & Chee, M. (1999). A novel method for determining linkage between DNA sequences: hybridization to paired probe arrays. *Nucl. Acids Res.* 27, 1485-1491.

Gunderson, K. L., Huang, X. C., Morris, M. S., Lipshutz, R. J., Lockhart, D. J. & Chee, M. S. (1998). Mutation detection by ligation to complete n-mer DNA arrays. *Genome Res.* 8, 1142-1153.

Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucl Acids Res* 22, 5456-5465

Guo, Z., Liu, Q. & Smith, L. M. (1997). Enhanced discrimination of single nucleotide polymorphisms by artificial mismatch hybridization *Nature Biotechnol* 15, 331-335.

Hacia, J. G. (1999). Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genet* 21, 42-47

Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P. & Collins, F. S. (1996) Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nature Genet*. 14, 441-447.

Hacia, J. G., Sun, B., Hunt, N., Edgemon, K., Mosbrook, D., Robbins, C., Fodor, S. P. A., Tagle, D. A. & Collins, F. S. (1998a). Strategies for mutational analysis of the large multiexon *ATM* gene using high-density oligonucleotide arrays. *Genome Res* 8, 1245-1258.

Hacia, J G., Woski, S. A., Fidanza, J., Edgemon, K., Hunt, N., McGall, G., Fodor, S. P. A. & Collins, F. S. (1998b). Enhanced high density oligonucleotide array-based sequence analysis using modified nucleoside triphosphates. *Nucl. Acids Res.* 26, 4975-4982.

Hasegawa, H., Ueda, M., Watanabe, M., Teramoto, T., Mukai, M. & Kitajima, M. (1995). K-*ras* gene mutations in early colorectal cancer . . . flat elevated *vs* polyp-forming cancer. *Oncogene.* 10, 1413-1416.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. & Holden, D. W. (1995). Simultaneous identification of bacterial virulence genes by negative selection *Science,* 269, 400-403.

Khanna, M., Park, P., Zirvi, M., Cao, W., Picon, A., Day, J., Paty, P. & Barany, F. (1999). Multiplex PCR/LDR for detection of K-*ras* mutations in primary colon tumors. *Oncogene,* 18, 27-38.

Kozal, M. J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T. C., Richman, D. D., Morris, D., Hubbell, E., Chee, M. & Gingeras, T. R. (1996). Extensive polymorphisms observed in HIV-I clade B protease gene using high-density oligonucleotide arrays. *Nature Med.* 2, 753-759.

Liu, B., Parsons, R., Papadopoulos, N., Nicolaides, N. C., Lynch, H. T., Watson, P., Jass, J. R., Dunlop, M., Wyllie, A., Peltomaki, P., de la Chapelle, A., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1996). Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nature Med.* 2, 169-174.

Lockley, A. K., Jones, C. G., Bruce, J. S., Franklin, S. J. & Bardsley, R. G. (1997). Colorimetric detection of immobilised PCR products generated on a solid support. *Nucl. Acids Res.* 25, 1313-1314.

Nikiforov, T. T., Rendle, R. B., Goelet, P., Rogers, Y. H., Kotewicz, M. L., Anderson, S., Trainor, G. L. & Knapp, M. R. (1994). Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucl. Acids Res.* 22, 4167-4175.

Nollau, P., Moser, C., Weinland, G. & Wagener, C. (1996). Detection of K-*ras* mutations in stools of patients with colorectal cancer by mutant-enriched PCR. *Int. J. Cancer,* 66, 332-336.

Parinov, S., Barsky, V., Yershov, G., Kirillov, E., Timofeev, E., Belgovskiy, A. & Mirzabekov, A. (1996). DNA sequencing by hybridization to microchip oct- and decanucleotides extended by stacked pentanucleotides. *Nucl. Acids Res.* 24, 2998-3004.

Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. & Syvanen, A. C. (1997). Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.* 7, 606-614.

Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holmes, C. P. & Fodor, S. P. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl Acad. Sci. USA,* 91, 5022-5026.

Perera, F. P. (1997). Environment and cancer: who are susceptible?. *Science,* 278, 1068-1073.

Sapolsky, R. J., Hsie, L., Berno, A., Ghandour, G., Mittman, M. & Fan, J.-B. (1999). High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays. *Genet. Anal.* 14, 187-192.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. & Davis, R. W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl Acad. Sci. USA,* 93, 10614-10619.

Shalon, D., Smith, S. J. & Brown, P. O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* 6, 639-645.

Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. (1996). Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* 14, 450-456.

Shumaker, J. M., Metspalu, A. & Caskey, C. T. (1996). Mutation detection by solid phase primer extension. *Hum. Mutat.* 7, 346-354.

Sidransky, D., Tokino, T., Hamilton, S. R., Kinzler, K. W., Levin, B., Frost, P. & Vogelstein, B. (1992). Identification of *ras* oncogene mutations in the stool of patients with curable colorectal tumors. *Science,* 256, 102-105.

Southern, E. M. (1996). DNA chips: analysing sequence by hybridization to oligonucleotides on a large scale. *Trends Genet.* 12, 110-115.

Southern, E., Mir, K. & Shchepinov, M. (1999). Molecular interactions on microarrays. *Nature Genet.* 21, 5-9.

Venitt, S. (1996). Mechanisms of spontaneous human cancers. *Environ. Health Perspect.* 3, 633-637.

Wang, D. G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T. & Hubbell, E., *et al.* (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science,* 280, 1077-1082.

Wu, S., Hoshino, B., Zhou, D. F. H., Liu, A., Aziz, D. C. & Barathur, R. R. (1994). Practical approaches to molecular screening of colon cancer. In *Early Detection of Cancer Molecular Markers* (Srivastava, S., Lippman, S. M., Hong, W. K. & Mulshine, J. L., eds), pp. 237-254, Futura Publishing Company, Inc., Armonk.

Yershov, G., Barsky, V., Belgovskiy, A., Kirillov, E., Kreindlin, E., Ivanov, I., Parinov, S., Guschin, D., Drobishev, A., Dubiley, S. & Mirzabekov, A. (1996). DNA analysis and diagnostics on oligonucleotide microchips. *Proc. Natl Acad. Sci. USA,* 93, 4913-4918.

Zhu, H., Cong, J.-P., Mamora, G., Gingeras, T. & Shenk, T. (1998). Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc. Natl Acad. Sci. USA,* 95, 14470-14475.

Exhibit 7:   Hashimoto et al., "Ligase Detection Reaction/Hybridization Assays Using Three-Dimensional Microfluidic Networks for the Detection of Low Abundant DNA Point Mutations," *Anal Chem* 77:3243–3255 (2005)

Ligase Detection Reaction/Hybridization Assays Using Three-Dimensional Microfluidic Networks for the Detection of Low-Abundant DNA Point Mutations

Masahiko Hashimoto,†,‡ Mateusz L. Hupert,†,‡ Michael C. Murphy,†,§ and Steven A. Soper*,†,‡,§

Center for Bio-Modular Multi-Scale Systems, Department of Chemistry, Department of Mechanical Engineering, Louisiana State University, Baton Rouge, Louisiana 70803

Yu-Wei Cheng and Francis Barany

Department of Microbiology, Joan and Sanford I. Weill Medical College of Cornell University, New York, New York 10021

We have fabricated a flow-through biochip assembly that consisted of two different microchips: (1) a polycarbonate (PC) chip for performing an allele-specific ligation detection reaction (LDR) and (2) a poly(methyl methacrylate) (PMMA) chip for the detection of the LDR products using an universal array platform. The operation of the device was demonstrated by detecting low-abundant DNA mutations in gene fragments (K-*ras*) that carry point mutations with high diagnostic value for colorectal cancers. The PC microchip was used for the LDR in a continuous-flow format, in which two primers (discriminating primer that carried the complement base to the mutation being interrogated and a common primer) that flanked the point mutation and were ligated only when the particular mutation was present in the genomic DNA. The miniaturized reactor architecture allowed enhanced reaction speed due to its high surface-to-volume ratio and efficient thermal management capabilities. A PMMA chip was employed as the microarray device, where zip code sequences (24-mers), which were complementary to sequences present on the target, were microprinted into fluidic channels embossed into the PMMA substrate. Microfluidic addressing of the array reduced the hybridization time significantly through enhanced mass transport to the surface-tethered zip code probes. The two microchips were assembled as a single integrated unit with a novel interconnect concept to produce the flow-through microfluidic biochip. A microgasket, fabricated from an elastomer poly(dimethylsiloxane) with a total volume of the interconnecting assembly of <200 nL, was used as the interconnect between the two chips to produce the three-dimensional microfluidic network. We successfully demonstrated the ability to detect one mutant DNA in 100 normal sequences with the biochip assembly. The LDR/

hybridization assay using the assembly performed the entire assay at a relatively fast processing speed: 6.5 min for on-chip LDR, 10 min for washing, and 2.6 min for fluorescence scanning (total processing time 19.1 min) and could screen multiple mutations simultaneously.

With completion of the sequencing of the human genome, new genes are being discovered at an accelerated pace as well as determining the function of these genes (functional genomics) and potential association of these genes and mutations within them to particular phenotypes (disease states). Efforts in functional genomics have produced an array of new diagnostic markers for clinical staging (prognosis), early detection, and predicting/monitoring the course of treatments for many genetic-based diseases. In most cases, a panel of mutations must be evaluated to obtain an accurate diagnosis/prognosis of that disease. For example, colorectal adenomas and cancers have been determined to possess point mutations in K-*ras* genes (19 different mutations), which can occur early in the development of the disease in nearly 30–50% of patients.[1-5] Most of these K-*ras* mutations are localized on codon 12 and to a lesser degree at codons 13 and 61.[6-11] Once acquired, K-*ras* mutations are conserved throughout the course

* To whom correspondence should be addressed. E-mail: chsopr@lsu.edu
‡ Center for Bio-Modular Multi-Scale Systems, Louisiana State University
† Department of Chemistry, Louisiana State University
§ Department of Mechanical Engineering, Louisiana State University

(1) Rothschild, C. B.; Brewer, C. S.; Loggie, B.; Beard, G. A.; Triscott, M. X. *J. Immunol. Methods* **1997**, *206*, 11–19.
(2) Otori, K.; Oda, Y.; Sugiyama, K.; Hasebe, T.; Mukai, K.; Fujii, T.; Tajiri, H.; Yoshida, S.; Fukushima, S.; Esumi, H. *Gut* **1997**, *40*, 660–663.
(3) Mulcahy, H. E.; Anker, P.; Chen, X.; Lefort, F.; Vasioukhin, V.; Lyautey, J.; Lederrey, C.; Struun, M.; Farthing, M. J. G. *Gastroenterology* **1996**, *110*, A563–A563.
(4) Chiang, J. M. *Cancer Lett.* **1998**, *126*, 179–185.
(5) Andersen, S. N.; Lovig, T.; Breivik, J.; Lund, E.; Gaudernack, G.; Meling, G. I.; Rognum, T. O. *Scand. J. Gastroenterol.* **1997**, *32*, 62–69.
(6) Bos, J. L. *Mutat. Res.* **1988**, *195*, 255–271.
(7) Breivik, J.; Meling, G. I.; Spurkland, A.; Rognum, T. O.; Gaudernack, G. *Br. J. Cancer* **1994**, *69*, 367–371.
(8) Capella, G.; Cronauer-Mitra, S.; Peinado, M. A.; Perucho, M. *Environ. Health Perspect.* **1991**, *93*, 125–131.
(9) Finkelstein, S. D.; Sayegh, R.; Bakker, A.; Swalsky, P. *Arch. Surg.* **1993**, *128*, 526–532.
(10) Forrester, K.; Almoguera, C.; Han, K.; Grizzle, W. E.; Perucho, M. *Nature* **1987**, *327*, 298–303.
(11) Losi, L.; Benhattar, J.; Costa, J. *Eur. J. Cancer* **1992**, *28A*, 1115–1120.

of tumor development For example, a single base substitution within codon 12 (GGT) in exon 1 of the K-ras gene may mutate to GAT, GCT, or GTT, being translated into the specific amino acid residues Asp, Ala, and Val, respectively. These amino acid residues play a critical role in GTP binding, and point mutations in these codons produce oncogenic p21 ras proteins that resist GTP hydrolysis and have constitutively active signaling functions.[14]

A major hurdle toward mutation detection is that the mutation of interest (mutant DNA) may be present in low copy numbers (minority) in a mixed population of higher copy number wild-type DNA (majority) Even at the primary tumor site for many cancers, the normal stromal cell content can be as high as 70% Therefore, if the mutation is found in only one of the two chromosomes of a tumor cell (heterozygous), the mutated DNA can be present as low as 15%[15] This number goes down precipitously if the sampling is done away from the primary tumor site. Thus, there is a need to develop technology that can identify accurately one or more low-abundant mutations at multiple adjacent, nearby, and distal loci in a large number of genes (multiplexed assays) with samples containing high stromal cell infiltration

One technique that can distinguish low-abundant mutant DNA from wild-type DNA is the ligase detection reaction (LDR) coupled to a primary PCR.[16-32] A conceptual schematic of the PCR/LDR technique is depicted in Figure 1 Following PCR amplification of the appropriate gene fragments, which contain sections of the gene(s) that possess the point mutations, the amplicon is mixed

with two LDR primers that flank the mutation of interest (common primer and discriminating primer) The discriminating primer contains a base at its 3'-end that coincides with the single-base mutation site If there is a mismatch, ligation of the two primers does not occur However, a perfect match results in a successful ligation of the two primers and produces a product that can be analyzed in a variety of fashions. The advantages of PCR/LDR are that it can be configured to do highly multiplexed assays and uses a thermally stable ligase enzyme to linearly amplify the LDR product

Recently, attention has focused on developing microfabricated devices for biological amplifications that require temperature regulation, such as PCR[33-48] and dideoxy cycle sequencing,[49] since they can offer lower thermal capacitance, require smaller amounts of reagents for the reaction, possess the potential for automation, and be integrated to subsequent processing steps configured on chips to improve automation and minimize sample contamination, which is extremely important to circumvent in clinical settings for early detection of a disease. During the past decade, a number of groups have designed chamber-type PCR chips, where a stationary PCR mixture in a confined space is alternately heated and cooled.[33-40] Alternatively, DNA amplification can be achieved in a microfluidic platform by shuttling a PCR cocktail in a microchannel repetitively through different isothermal zones using a continuous-flow (CF) format.[41-48] The CFPCR approach can be conducted at relatively high speeds since it is not necessary to heat and cool the large thermal mass associated with the amplification chamber repeatedly. We have developed a unique spiral microchannel with 20 loops hot-embossed into polycarbonate (PC) for rapid CFPCR [46-48] Ultrafast PCR was demonstrated with the speed of the reaction determined primarily by enzyme kinetics (AmpliTaq polymerase) using a flow velocity of 15 mm s^{-1} resulting in successful amplification of a 500-bp fragment in 1.7 min at a cycle rate of 5 s cycle^{-1} [46]

(12) Smith, A. J.; Stern, H. S.; Penner, M.; Kazy, H.; Mitri, A.; Bapat, B. V.; Gallinger, S. *Cancer Res* 1994, 54, 5527–5530
(13) Vogelstein, B.; Fearon, E. R.; Hamilton, S. R.; Kern, S. E.; Preisinger, A. C.; Leppert, M.; Nakamura, Y.; White, R.; Smits, A. M. M.; Bos, J. L. *New Engl J Med* 1988, 319, 525–532
(14) Barbacid, M *Annu Rev Biochem* 1987, 56, 779–827
(15) Gerry, N. P.; Witowski, N. E.; Day, J.; Hammer, R. P.; Barany, G.; Barany, F. *J Mol Biol* 1999, 292, 251–262
(16) Barany, F *Proc Natl Acad Sci USA* 1991, 88, 189–193
(17) Wei, Q.; Barany, F.; Wilson, V. L. *Mol Biol Cell* 1992, 3, A22–A22
(18) Wiedmann, M.; Barany, F.; Batt, C. A *Appl Environ Microbiol* 1993, 59, 2743–2745
(19) Wiedmann, M.; Wilson, W. J.; Czajka, J.; Luo, J. Y.; Barany, F.; Batt, C. A *PCR Methods Appl* 1994, 3, S51–S64
(20) Day, D. J.; Speiser, P. W.; White, P. C.; Barany, F. *Genomics* 1995, 29, 152–162.
(21) Luo, J. Y.; Bergstrom, D. E.; Barany, F *Nucleic Acids Res* 1996, 24, 3071–3078
(22) Luo, J. Y.; Barany, F *Nucleic Acids Res* 1996, 24, 3079–3085
(23) Khanna, M.; Cao, W. G.; Zirvi, M.; Paty, P.; Barany, F. *Clin. Biochem.* 1993, 32, 287–290
(24) Khanna, M.; Park, P.; Zirvi, M.; Cao, W. G.; Picon, A.; Day, J.; Paty, P.; Barany, F. *Oncogene* 1999, 18, 27–38
(25) Tong, J.; Cao, W. G.; Barany, F *Nucleic Acids Res* 1999, 27, 788–794
(26) Favis, R.; Day, J. P.; Gerry, N. P.; Phelan, C.; Narod, S.; Barany, F *Nat Biotechnol* 2000, 18, 561–564
(27) Huang, J.; Kirk, B.; Favis, R.; Soussi, T.; Paty, P.; Cao, W.; Barany, F *Oncogene* 2002, 21, 1909–1921.
(28) Wang, Y.; Vaidya, B.; Farquar, H. D.; Stryjewski, W.; Hammer, R. P.; McCarley, R. L.; Soper, S. A.; Cheng, W.; Barany, F *Anal Chem* 2003, 75, 1130–1140
(29) Wabuyele, M. B.; Farquar, H.; Stryjewski, W.; Hammer, R. P.; Soper, S. A.; Cheng, Y.-W.; Barany, F. *J Am Chem Soc* 2003, 125, 6937–6945
(30) Nathanson, D. R.; Culliford, A. T., IV; Shia, J.; Chen, B.; D'Alessio, M.; Zeng, Z. S.; Nash, G. M.; Gerald, W.; Barany, F.; Paty, P. B *Int J Cancer* 2003, 105, 796–802
(31) Thomas, G.; Sinville, R.; Sutton, S.; Farquar, H.; Hammer, R. P.; Soper, S. A.; Cheng, Y. W. Barany, F. *Electrophoresis* 2004, 25, 1668–1677
(32) Favis, R.; Huang, J.; Gerry, N. P.; Culliford, A.; Paty, P.; Soussi, T.; Barany, F *Hum. Mutat* 2004, 24, 63–75

(33) Lagally, E. T.; Medintz, I.; Mathies, R. A *Anal Chem*, 2001, 73, 565–570
(34) Lagally, E. T.; Emrich, C. A.; Mathies, R. A. *Lab Chip* 2001, 1, 102–107
(35) Giordano, B. C.; Ferrance, J.; Swedberg, S.; Huhmer, A. F. R.; Landers, J. P *Anal. Biochem.* 2001, 291, 124–132
(36) Khandurina, J.; McKnight, T. E.; Jacobson, S. C.; Waters, L. C.; Foote, R. S.; Ramsey, J. M *Anal Chem* 2000, 72, 2995–3000
(37) Koh, C. G.; Tan, W.; Zhao, M.; Ricco, A. J.; Fan, Z. H. *Anal Chem* 2003, 75, 4591–4598
(38) Burns, M. A.; Mastrangelo, C. H.; Sammarco, T. S.; Man, F. P.; Webster, J. R.; Johnsons, B. N.; Foerster, B.; Jones, D.; Fields, Y.; Kaiser, A. R.; Burke, D. T. *Proc. Natl Acad Sci USA* 1996, 93, 5556–5561.
(39) Liu, Y.; Rauch, C. B.; Stevens, R. L.; Lenigk, R.; Yang, J.; Rhine, D. B.; Grodzinski, P *Anal. Chem.* 2002, 74, 3063–3070.
(40) Trau, D.; Lee, T. M. H.; Lao, A. I. K.; Lenigk, R.; Hsing, I.-M.; Ip, N. Y.; Carles, M. C.; Sucher, N. J *Anal Chem* 2002, 74, 3168–3173
(41) Liu, J.; Enzelberger, M.; Quake, S. *Electrophoresis* 2002, 23, 1531–1536.
(42) Kopp, M. U.; De Mello, A. J.; Manz, A. *Science* 1998, 280, 1046–1048.
(43) Curcio, M.; Roeraade, J. *Anal Chem* 2003, 75, 1–7
(44) Obeid, P. J.; Christopoulos, T. K.; Crabtree, H. J.; Backhouse, C. J. *Anal Chem* 2003, 75, 288–295.
(45) Park, N.; Kim, S.; Hahn, J. H. *Anal. Chem* 2003, 75, 6029–6033
(46) Hashimoto, M.; Chen, P.-C.; Mitchell, M. W.; Nikitopoulos, D. E.; Soper, S. A.; Murphy, M. C *Lab Chip* 2004, 4, 638–645
(47) Liu, X.; Mitchell, M. W.; Bejat, Y.; Nikitopoulos, D. E.; Soper, S. A.; Murphy, M. C.; *Microelectromech Syst*, in revision
(48) Bejat, Y.; Mitchell, M. W.; Nikitopoulos, D. E.; Liu, X.; Soper, S. A.; Murphy, M. C. *J Microelectromech Syst*, in revision.
(49) Oda, R. P.; Strausbauch, M. A.; Huhmer, A. F. R.; Borson, N.; Jurrens, S. R.; Craighead, J.; Wettstein, P. J.; Eckloff, B.; Kline, B.; Landers, J. P *Anal Chem* 1998, 70, 4361–4368
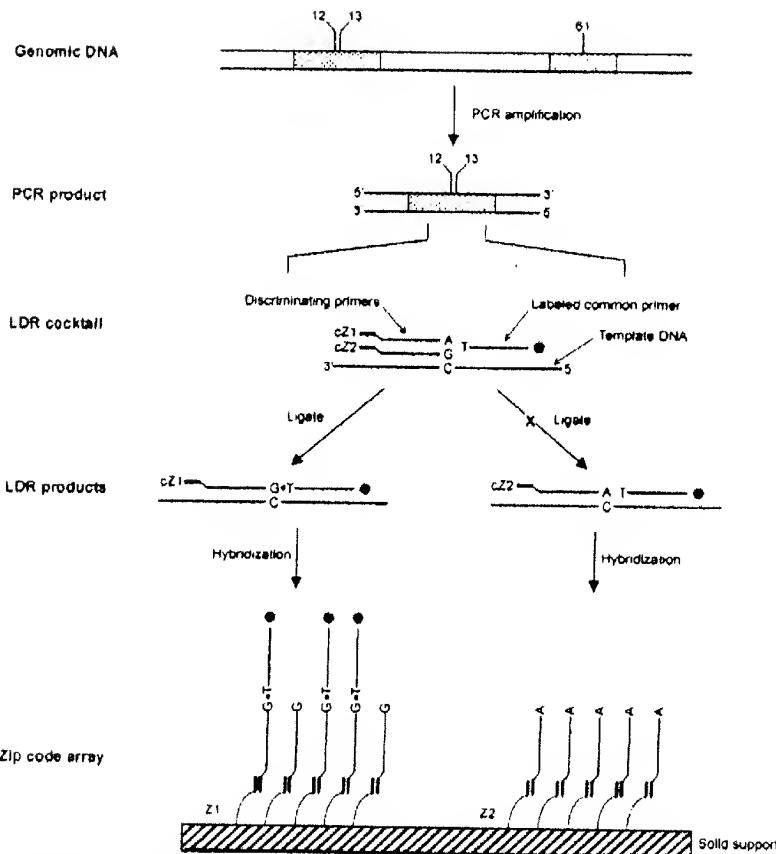
Figure 1. Conceptual schematic of the PCR/LDR/universal array assay

DNA microarrays can be configured to detect sequence variations at many different sites simultaneously for potential diagnostic applications.[50-58] Recently, universal zip code arrays have been developed for monitoring products generated from allele-specific reactions, such as LDR.[15,46,58] The array format (see Figure 1) uses small probes that serve as zip codes (24-mers with similar T_{m} values) that contain unique sequences not found in the sample DNA template. The PCR/LDR uses discriminating and

common primers similar to that described above. However, the allele-specific primer contains on its 5'-end a zip code complement that directs the ligation product to a particular address on the array. The common primer contains a fluorescent dye on its 3'-end. If the mutation is present, LDR ligates the two primers together and generates a fluorescence signal at the appropriate location of the array. The attractive feature of these universal arrays is that they can be used to detect a variety of mutations by simply appending the correct zip code complement sequence to the discriminating primer used in the LDR step.

Several reports have discussed merging microarrays with microfluidics. The attractive features of this marriage include the reduced amount of sample required to address each element of the array, the enhanced mass transport to the array surface reducing analysis time, the ability to monitor several samples simultaneously using multichannel chips, the ability to integrate several preprocessing steps into the microfluidic device, and the closed architecture of the microfluidics reducing the potential for

(50) Consolandi, C.; Castiglioni, B.; Bordoni, R.; Busti, F.; Battaglia, C.; De Bellis, G. *Minerva Biotecnol.* **2001**, *13*, 261–268.
(51) De Benedetti, V. M. G.; Biglia, N.; Sismondi, P.; De Bortoli, M. *Int. J. Biol. Markers* **2000**, *15*, 1–9.
(52) Dimitrijevic, B. *Jugosl. Med. Biohem.* **2001**, *20*, 65–71.
(53) Havw, I.; Campbell, I. G. *Mol. Cell. Endocrinol.* **2002**, *191*, 121–126.
(54) Kolchinsky, A.; Mirzabekov, A. *Hum. Mutat.* **2002**, *19*, 343–360.
(55) Lee, P.; Hudson, T. J. *M. S-Med. Sci.* **2000** *16*, 43–49.
(56) Petrik, J. *Vox Sang.* **2001**, *80*, 1–11.
(57) Tawata, M.; Aida, K.; Onaya, T. *Comb. Chem. High Throughput Screening* **2000**, *3*, 1–9.
(58) Zakhrabekova, S.; Kannangara, C. G.; von Wettstein, D.; Hansson, M. *Plant Physiol. Biochem.* **2002** *40*, 189–197.

Table 1. Sequences of Oligonucleotides Used in the PCR/LDR/Array Hybridization Assays

oligos	sequences (5′→3′)	T_m (°C)
K-ras exon 1 forward	TTAAAAGGTACTGGTGGAGTATTTGATA	56.0
K-ras exon 1 reverse	AAAATGGTCAGAGAAACCTTTATCTGT	56.8
K-ras c12 Com 2	p[a]TGGCGTAGGCAAGAGTGCCT-DY782[b]	62.5
cZip1-K-ras c12 2W[c]G	[c]GCTGAGGTCGATGCTGAGGTCGCAAAACTTGTGGTAGTTGGAGCTGG	74.6
cZip3-K-ras c12 2D	[c]GCTGCGATCGATGGTCAGGTGCTGAAACTTGTGGTAGTTGGAGCTGA	74.3
cZip5-K-ras c12.2A	[c]GCTGTACCCGATCGCAAGGTGGTCAAACTTGTGGTAGTTGGAGCTGC	74.7
cZip11-K-ras c12 2V	[c]CGCAAGGTAGGTGCTGTACCCGCAAAACTTGTGGTAGTTGGA GCTGT	74.5
zip code 1	TGCGACCTCAGCATCGACCTCAGC-spacer-NH₂[d]	66.4
zip code 3	CAGCACCTGACCATCGATCGCAGC-spacer-NH₂[d]	65.5
zip code 5	GACCACCTTGCGATCGGGTACAGC-spacer-NH₂[d]	65.1
zip code 11	TGCGGGTACAGCACCTACCTTGCG-spacer-NH₂[d]	66.7

[a] p, phosphorylated. [b] DY782, $\lambda_{ex} = 782$ nm; $\lambda_{em} = 800$ nm. [c] The boldface sequences are complementary to the sequences of zip code probes. [d] Spacer-NH₂, PO₄(CH₂CH₂O)₆PO₄CH₂CH(CH₂OH)(CH₂)₄NH₂

sample contamination. Anderson and co-workers developed a highly integrated monolithic device that automatically carried out a complex series of molecular processes on multiple samples.[59] The device could manipulate over 10 reagents in more than 60 sequential operations and was tested for the detection of mutations in a 1.6-kb region of the HIV genome from serum samples containing as few as 500 copies of the target RNA. Check et al developed three-dimensional flow-through arrays in glass chips by fabricating channels (10-μm i d.) that were 39 pl. in volume and 5 μm in radius with the DNA probes printed on the channel walls.[60] Liu and co-workers fabricated a plastic (PC) chip that integrated microfluidic mixers, valves, pumps, channels, chambers, heaters, and DNA microarray sensors.[61] The hybridization probes were tethered to the surface of gold electrodes via self-assembly to detect electrochemical signals generated from hybridized target DNA. The analysis required, from loading sample and different reagents into storage chambers to obtaining the hybridization results, nearly 3.5 h of processing time. One potential problem associated with the use of PC as the microarray platform is that it generates a significant amount of autofluorescence, which can potentially degrade detection limits when fluorescence readout is used.[62] In addition, PC can produce large levels of nonspecific adsorption in its native state.[63]

In this paper, we report the development of a polymer flow-through biochip assembly that consists of a continuous-flow LDR (CFLDR) microchip and a microarray chip for the detection of low-abundant mutations in genomic DNA. We chose PC as the material for CFLDR due to its high glass transition temperature (145–148 °C) allowing it to withstand the sustained high operating temperatures associated with LDR (~95 °C for thermal denaturation). On the other hand, the material of choice for the universal array chip was poly(methyl methacrylate) (PMMA) because PMMA has significantly lower amounts of autofluorescence compared to PC as well as minimal nonspecific adsorption artifacts

(59) Anderson, R. C., Su, X., Bogdan, G. J., Fenton, J. *Nucleic Acids Res* 2000, 28, e60.

(60) Cheek, B. J., Steel, A. B., Torres, M. P., Yu, Y.-Y., Yang, Y. *Anal. Chem.* 2001, 73, 5777–5783.

(61) Liu, R. H., Yang, J., Lenigk, R., Bonanno, J., Grodzinski, P. *Anal. Chem.* 2004, 76, 1824–1831.

(62) Wabuyele, Musundi B., Ford, Sean M., Stryjewski, Wieslaw, Barrow, James, Soper, Steven A. *Electrophoresis* 2001, 22, 3939–3948.

(63) Xu, Y., Vaidya, B., Patel, A. B., Ford, S. M., McCarley, R. L., Soper, S. A. *Anal. Chem.* 2003, 75, 2975–2984.

between PMMA and single-stranded DNAs. In addition, we have outlined robust procedures for covalently attaching oligonucleotide probes to the surface of PMMA at high concentrations.[28,64,65]

The need to integrate specific functions onto different chip materials through a three-dimensional microfluidic network led us to develop a simple, low-volume interconnect technology that provided robust leakage-free connection with minimum dead volume between the chips. We utilized an elastomer, poly-(dimethylsiloxane) (PDMS), as an O-ring gasket between the chips in order to connect laser-drilled microchannels between the two chips. In this work, a PCR/LDR/hybridization assay was carried out on K-ras genes to detect the presence of point mutations possessing clinical relevance for diagnosing colorectal cancers. The PC chip used for LDR employed a CF format with isothermal zones.[46-48] Using presynthesized oligonucleotide probes, the array containing the zip code probes was microprinted into microfluidic channels hot-embossed in a PMMA chip. Following production of the LDR and array microchips, the array was embedded into a machine-milled pocket in the LDR chip with the PDMS gasket for the construction of the three-dimensional microfluidic network.

EXPERIMENTAL SECTION

Reagents and Materials. PC and PMMA used as the microfluidic chip substrates were purchased from GoodFellow (Berwyn, PA). Chemicals used for PMMA surface modification and hybridization assays, including n-butyllithium, ethylenediamine, 50 wt % glutardialdehyde, sodium borohydride, sodium cyanoborohydride (5.0 M solution in aqueous ~1 M sodium hydroxide), and 20× sodium chloride–sodium phosphate–EDTA (SSPE) buffer, were purchased from Aldrich Chemical (Milwaukee, WI) and used as received. A 10% SDS stock solution, which was used for posthybridization washing, was received from Ambion (Austin, TX). Oligonucleotide probes and primers were obtained from two different sources, Integrated DNA Technologies (Coralville, IA) and Synthegen (Houston, TX). Their sequences and melting temperatures (T_m) are listed in Table 1. *Thermus thermophilus* (Tth) and *Thermus aquaticus* (Taq) DNA ligases were

(64) Henry, A. C., Tutt, T. J., Galloway, M., Davidson, Y. Y., McWhorter, C. S., Soper, S. A., McCarley, R. L. *Anal. Chem.* 2000, 72, 5331–5337.

(65) Waddell, E., Wang, Y., Stryjewski, W., McWhorter, S., Henry, A. C., Evans, D., McCarley, R. L., Soper, S. A. *Anal. Chem.* 2000, 72, 5907–5917.
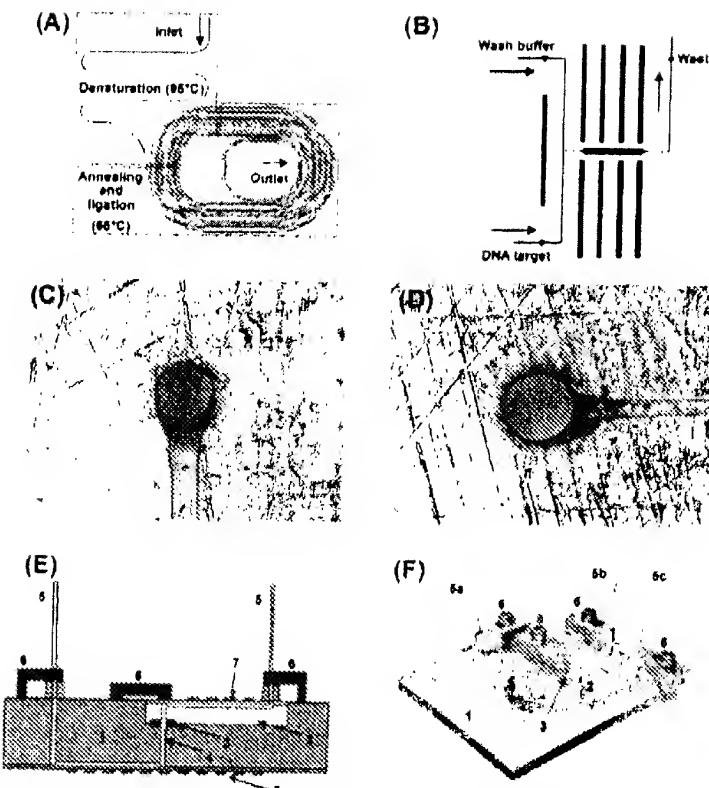
Figure 2. Schematic representation of the LDR/zip code array biochip assembly. (A) PC CFLDR chip and (B) PMMA zip code array microchip. The spiral channel was 50 μm in width, 150 μm in depth, and ~1.6 m in length and consisted of a 20-turn loop. The hybridization chamber was 500 μm in width, 50 μm in depth, and ~6.7 mm in length. (C) 100- and (D) 270-μm holes laser-drilled into the LDR chip at the end of a 50-μm-i.d. microchannel. (E) Side-view diagram and (F) top-view picture of the assembled microchip: 1, PC LDR microfluidic chip; 2, PMMA zip code array microfluidic chip; 3, PDMS gasket; 4, laser-drilled interconnecting microchannel; 5, fused-silica capillaries (a, inlet to LDR microchip; b, inlet of wash buffer to zip code array chip; c, drain); 6, clamps; 7, Kapton film heaters. The channels were filled with ink and heaters were removed for clarity.

obtained from ABgene (Rochester, NY) and New England Biolabs (Beverly, MA), respectively.

Chip Fabrication. Microchips were fabricated using methods previously reported in our group.[66,67] Briefly, the procedure involved fabricating a metal molding die by LiGA, which consisted of raised Ni microstructures electroplated from a Ni sulfamate bath onto a stainless steel base plate. The topographical layout of the CFLDR chip is depicted in Figure 2A. The spiral channel (~1.6 m long), which consisted of 20 loops, was 50 μm in width, 150 μm in depth, and a 250 μm interchannel spacing. The topology of the microarray chip is shown in Figure 2B. Fluid access channels (DNA sample and wash buffers) were 100 μm in width and 50

(66) Ford, S. M.; Davies, J.; Kar, B.; Qi, S. D.; McWhorter, S.; Soper, S. A.; Malek, C. K. *J. Biomech. Eng.-Trans. ASME* **1999**, *121*, 13-21.

(67) Galloway, M.; Stryjewski, W.; Henry, A.; Ford, S. M.; Llopis, S.; McCarley, R. L.; Soper, S. A. *Anal. Chem.* **2002**, *74*, 2407-2415.

μm in depth. Both channels merged into one common channel and emptied into the hybridization chamber, which measured 500 μm in width, 50 μm in depth, and 6.7 mm in length (total volume, 168 nL). Replicates from the molding die were hot-embossed into the plastic substrates. The embossing system consisted of a PHI Precision Press model number TS-21-H-C (4A)-5 (City of Industry, CA). A vacuum chamber was installed into this press to remove air (pressure, <0.1 bar) to minimize replication errors. During embossing, the molding die for the spiral microchannel was heated to 190 °C and pressed into the PC wafer with a force of 850 lb for 5 min. The microchannel pattern for the array chip was hot-embossed into a PMMA plate at 155 °C and 1000 lb for 4 min. After hot-embossing, the press was opened and the polymer part removed and cooled to room temperature.

Drilling of interconnecting microchannels at the ends of the hot-embossed microchannels was performed using a KrF laser

system (RapidX 1000 Series, Resonetics Inc., Nashua, NH) with a laser fluence at the workpiece of ~10 J/cm² and a repetition rate of 50 Hz. Micrographs of the laser-drilled holes (100- and 270-μm i.d.) are shown in Figure 2C and D.

Milling of the pocket on the backside of the LDR microchip for the loading of the microarray chip was performed using a Kern MMP 2522 micromilling machine (Kern Micro- und Feinwerktechnik GmbH & Co KG, Murnau, Germany). Milling was performed using a 500-μm milling bit at 40 000 rpm to provide a smooth finish to the surface, which was necessary to obtain a good seal between the microfluidic chips.

The embossed PMMA array chips were assembled by heat annealing a cover plate made from the same material to the substrate at 107 °C for 20 min. The cover plates (500 μm thick) and substrates (2 mm thick) were clamped together and placed in a convection oven. Sealing of the PC spiral microchannel with a cover plate was more critical due to the elaborate pattern of the microchannel topography and the higher glass transition temperature of PC. The embossed substrate (4.5-mm thickness) and the cover plate (250-μm thickness) were introduced into the embossing machine, and the assembly was heated to 160 °C for 15 min under vacuum for a tight seal of PC to PC.

Surface Modification and Array Preparation. The hot-embossed PMMA microchannels were treated (prior to chip assembly) with a procedure that was slightly modified from our previously published method.[38,64,65] A N-lithioethylenediame solution was prepared by dissolving n-butyllithium in ethylenediamine at an equal volume ratio (50:50) and magnetically agitating the solution until it turned dark purple. PMMA microchannels were aminated by immersing them in the N-lithioethylenediamine solution at room temperature for 5 min under nitrogen atmosphere. After amination was complete, the chips were thoroughly rinsed with doubly distilled H_2O. Next, the aminated chips were soaked for 2 h in a 5% glutardialdehyde (cross-linking agent) solution with 50 μM sodium cyanoborohydride in phosphate buffer (0.5 M, pH 6.4). Zip code oligonucleotide probes 1, 3, 5, and 11 (see Table 1 for sequences) were dissolved separately in 0.2 M phosphate buffer (pH 8.3) to a final concentration of 10 μM. The arrays were printed onto the activated PMMA. The volume of probe (zip code sequence) deposited was 100 nL, which resulted in a spot of ~400 μm in diameter. A spot for each probe was placed along the length of the hybridization chamber (linear array). After attachment of the oligonucleotide probes, excess surface aldehydes were capped with 0.25% (w/v) sodium borohydride solution in phosphate buffer (0.1 M, pH 6.1). Following deposition of the four probes and capping, the chip was assembled by placing a cover plate over the microfluidic channels, clamping the assembly together, and thermally annealing using the procedure described above.

Near-IR Laser Scanner. Near-IR fluorescence measurements were made from the arrays with a device built in-house that has been described previously.[6] Briefly, the near-IR scanner consisted of a diode laser (PicoQuant GmbH, model 800, Berlin, Germany), counting electronics (PicoQuant GmbH, model SPC 430), and a single-photon avalanche diode (SPAD, EG&G Optoelectronics, model SPCM-PQ, Vaudreuil, Canada). The components were mounted with the aid of a mounting cube and lens tubes purchased from ThorLabs (Newton, NJ) and configured in an epi-

illumination format. The diode laser operated at a wavelength of 780 nm with 7.5 mW of average power. An integrated optics set was provided with the laser to produce an elliptically shaped collimated output. The laser excitation beam was passed through a 780-nm line filter (Omega Optical, 780DF10, Brattleboro, VT), reflected by a dichroic mirror (Omega Optical, 795DRLP) and focused onto the array surface using a 40× high numerical aperture (NA) microscope objective (Nikon, Natick, MA, NA = 0.85). The $1/e^2$ spot size of the excitation beam was measured to be 3 μm (minor axis) by 5 μm (major axis). The fluorescence excitation was collected by the same microscope objective, transmitted through the dichroic, a circular aperture set at 2.0 mm (clear diameter), and finally through a filter stack consisting of a long-pass filter (cut-on wavelength, 830 nm; Newport Corp., Irvine, CA) and a band-pass filter centered at 825 nm (825RDF30, Omega Optical). After passing through the filters, the fluorescence was sent through a condensing lens (01LAG111/076, Melles Griot) and focused onto a single photon avalanche diode.

The entire fluorescence detector was mounted on an X/Y microtranslational stage. Two bipolar stepper motors interfaced to a PC using STP-100 stepper motor controller boards obtained from Pontech, Inc. (Upland, CA) drove the X and Y directions of the microtranslational stages. Each STP-100 was equipped with the RS-485 interface allowing full duplex, multidrop communication with the host computer. A PP232-485I interface (Pontech, Inc.) was used to convert RS485 into the PC's RS232 protocol. Hall sensors were used to monitor the travel limits of the X/Y stages. The step resolution of these stages was either 25.4 or 50.8 μm with a maximum scan range of 4 cm in both the X and Y coordinates. The scanner operated by taking a single step and then acquiring the fluorescence data for a software-selectable integration period (10 ms-10 s).

The data acquisition software was written in Visual Basic and consisted of several control and data acquisition functions such as recording the position of the scanning head, streaming data to the hard drive, and providing real-time visualization of the acquired images. The size of each data file was determined by the number of pixels included in the image file (set by the stepping resolution and area imaged) with four bytes representing the intensity at each image pixel.

Extraction of DNA from Cell Lines. Genomic DNA was extracted from cell lines of known K-*ras* genotype (HT29, wild-type; SW1116, G12A; LS180, G12D; SW620, G12V).[24] Cell lines were grown in RPMI culture media with 10% bovine serum. Harvested cells (~1 × 10⁷) were resuspended in DNA extraction buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM EDTA, pH 8.0) containing 0.5% SDS and 200 μg/mL proteinase K and incubated at 37 °C for 4 h. Thirty percent (v/v) of 6 M NaCl was added to the mixture, and the samples were centrifuged. DNA was precipitated from the supernatant with three volumes of EtOH, washed with 70% EtOH, and resuspended in TE buffer (10 mM Tris-HCl, pH 7.2, 2 mM EDTA, pH 8.0).

PCR Amplification of Genomic DNA. PCR amplifications were carried out using a commercial thermal cycling machine (Techne, Burlington, NJ) in 50 μl of 10 mM Tris-HCl buffer (pH 8.3) containing 50 mM KCl, 1.5 mM MgCl₂, 200 μM dNTPs, 200 nM forward and reverse primers, and between 1 and 50 ng of genomic DNA extracted from the cell lines. The primers used

were as follows: forward = 5′ TTA AAA GGT ACT GGT GGA GTA TTT GAT A 3′: reverse = 5′ AAA ATG GTC AGA GAA ACC TTT ATC TGT 3′. After a 2-min denaturation step, 5.0 units of Amplitaq DNA polymerase (Perkin-Elmer, Norwalk, CT) was added under hot start conditions and amplification achieved by thermal cycling for 35–40 cycles at 95 °C for 30 s, 60 °C for 2 min. and a final extension at 72 °C for 3 min. PCR products were quantified by absorbance at 260 nm and stored at –20 °C until required for the LDR assays.

LDR and Electrophoresis of LDR Products. Reference LDRs were executed in a total volume of 100 µL in 0.2-mL polypropylene microtubes using a commercial thermal cycling machine (Genius Series 96-well Thermal Cycler, Techne, Minneapolis, MN). The reaction cocktail typically employed in this work consisted of 20 mM Tris-HCl (pH 8.3), 25 mM KCl, 10 mM MgCl₂, 0.5 mM NAD⁺ (nicotinic adenine dinucleotide, a cofactor for ligase enzyme), 0.01% Triton X-100, 30 nM discriminating primers (7.5 nM each), 30 nM com-2 fluorescently labeled primer (see Table 1 for sequences of primers), mixture of PCR products (wild-type and mutant DNA), and 2 units/µL of ligase enzyme. The concentration ratio of the mutant-to-wild-type DNA was adjusted from 0.1 (mutant/wild-type, control) to 1:1000. It is known that incorporation of bovine serum albumin (BSA; ~0.5 µg/µL) into a reaction mixture is essential in order to avoid deactivation of the bioenzyme due to its nonspecific adsorption to PC surfaces.[?],[??],[??] Our studies indicated that this procedure also could effectively prevent the majority of nonspecific adsorption of ligase enzyme to PC used as our substrate material for CFLDR.[?] The LDR mix was preheated to 94 °C for 2 min and then subjected to 20 LDR thermal cycles using the following temperatures: 94 °C for 30 s; 65 °C for 15 s–2 min. To test the fidelity and yield of the LDR reaction, slab gel electrophoresis was run on an aliquot of each reaction (1 µL of LDR product was mixed with 2 µL of loading dye and then 1 µL of that mixture was loaded into an individual well of a slab gel).

Electrophoresis was accomplished using a 5.5% (w/v) crosslinked polyacrylamide gel (Li-COR Biotechnology, Lincoln, NE). The gel was polymerized between two borofloat glass plates (21 cm × 25 cm) and placed in the Global IR² DNA analysis system (Li-COR Biotechnology). The electrophoresis was typically run at ~1500 V for 2 h. The fluorescence bands were integrated over each separation lane with ImageQuant software (Amersham Biosciences, Piscataway, NJ)

Operational Protocol of the Assembled Biochip. Panels E and F in Figure 2 show a schematic diagram and picture of the flow-through biochip assembly. The construction of the biochip was conducted in the following manner. A PDMS O-ring (~1 mm i.d., ~3-mm o.d., 500 µm thick) (3) used as a gasket was placed on top of the laser-drilled microchannel (100-µm diameter) (4) of the LDR chip (1). The PMMA array microchip (2) was next carefully loaded into the mechanically milled pocket of the LDR chip so that the two microchannels (4) and the PDMS gasket (3) aligned properly in order to provide the flow-through microfluidic network. The array chip was tightly pressed onto the LDR chip using mechanical clamps (6) to seal the two chips between the PDMS gasket, allowing a leak-free connection, which was deter-

mined by flowing fluorescent dye (fluorescein) through the interconnect (data not shown). Clamping pressed the PDMS gasket to ~400-µm thickness and ~100-µm-i.d. through-hole, leading to a total volume of the interconnect assembly (laser-drilled microchannels and gasket) of <200 nl. (~1.6% of CFLDR reactor volume). Film resistance heaters (KHLV-101/10, Omega Engineering, Inc., Stamford, CT) [7] were attached to the cover plate of the LDR chip and the array chip for thermal control of hybridization reactions. Three capillary tubes (75-µm i.d.; 363-µm o.d.; 18 cm long, Polymicro Technologies, Phenix, AZ) [5] were then affixed to the chips to aid in loading.

A syringe pump (Pico Plus, Harvard Apparatus, Holliston, MA) was used to drive the LDR mixture through the spiral microchannel via a capillary tube (5a). A glass syringe (Hamilton, Reno, NV) with a syringe-to-capillary adapter was used to make the connection between the pump and the microfluidic assembly. Temperatures were maintained during operation using the heaters under closed-loop PID control (CN77R340, Omega Engineering, Inc.) Temperature feedback was supplied through type-K thermocouples (5TC-TT-K-36-36, Omega Engineering, Inc.) mounted between the cover plates and heaters. The arrangement of temperature zones on the spiral channel (95 °C for denaturing and 65 °C for annealing and ligation) is depicted in Figure 2A. The resultant LDR products were directly pumped into the hybridization chamber via the laser-drilled microchannels (4) and the PDMS gasket (3) and subjected to hybridization to the surface-tethered zip code probes at 55 °C. The hybridization chamber was flushed with a wash buffer (2× SSPE-0.1% SDS) at a volumetric flow rate of 4.4 µL/min, and the array chip was finally imaged with the near-IR laser-induced fluorescence scanner.

RESULTS AND DISCUSSION

Effects of Flow Rate on LDR Product Yield. The speed of thermal cycling is usually determined by thermal conduction and mass of the polypropylene containers and the heating block. For example, commercial PCR machines are based on temperature-controlled metal block holding tubes containing the PCR cocktail that is thermally cycled during the PCR. Standard protocols for 30 thermal cycles can require >2 h of processing time with a large fraction of that time required to bring the PCR cocktail to the set temperature due to the need for bringing the large metal block to the cycle equilibrium temperature and to transfer heat to the cocktail through the microfuge tubes. Therefore, the cycle time is set by the thermal capacitance of the metal block and the heat transfer through the plastic microfuge tubes. Theoretically, thermal cycling can be carried out much more rapidly if the sample volume is small, the container wall is thin, there is reduction in the thermal mass required to be heated/cooled, or the surface-to-volume ratio of the sample reaction chamber is high. Hence, the use of a microreaction chamber operated in a continuous-flow format is very attractive due to the high surface-to-volume ratio of the device and the fact that the entire system is brought to thermal equilibrium prior to operation (i.e., except for the small fluid packets that are transferred from isothermal zone to another zone). A feature of the spiral microchannel in this work is its high surface-to-volume ratio (SVR, 623 mm², volume, 12 µl; SVR = 51.9 mm⁻¹) This is much higher than a conventional

(66) Wittwer, C. T.; Fillmore, G. C.; Garling, D. J. *Anal. Biochem.* 1990, *186*, 328–331.

plastic reaction tube (77 mm², 50 μL; SVR = 1.5 mm⁻¹).[64] Another advantage of the continuous-flow format employed in this work is that it is not necessary to heat or cool the amplification vessel repeatedly, allowing thermal cycling to be carried out at relatively high speeds

It is known that denaturing and renaturation are almost instantaneous (less than 1 s).[64] Therefore, LDR cycle times should ultimately be limited by the ligation time, which is determined primarily by enzyme kinetics. Figure 3 shows the influence of on-chip ligation time, which is determined by the dwell time of the LDR cocktail within the 65 °C zone (refer to Figure 2A) on the LDR product yield. Ligation times of 2, 1, 0.5, and 0.25 min were investigated using volumetric flow rates of 0.22, 0.45, 0.9, and 1.8 μL/min, respectively, to control the dwell time within the ligation zone It should be noted that a sufficient dwell time was generated for denaturation even when the flow rate was 1.8 μL/min (denaturation time at the innermost loop <2.5 s). The resultant LDR products were collected into microfuge tubes from the outlet of the spiral channel and analyzed using a 5.5% cross-linked polyacrylamide gel. Fluorescence from the product was imaged by the gel electrophoresis instrument (Figure 3A) and the resultant band integrated over each separation lane with ImageQuant software for quantification (Figure 3B). Using the ligation times employed in this series of measurements, the longer the ligation time, the larger the amount of product that was obtained. For example, the product yield with a 2-min ligation time was 5× larger than that seen with a 15-s ligation time. However, the 2-min ligation time required an extra processing time of 45 5 min compared to a 15-s ligation time (52 min with a 2-min ligation and 6.5 min with a 15-s ligation time for a total of 20 LDR cycles).

Comparable amounts of product for the microchip and the reference thermal cycler could be obtained using a 4-min ligation time (data not shown). However, the difference in product yield between on-chip LDR and the reference LDR became larger as the ligation time was reduced (see Figure 3C). The on-chip LDR product yields were 74 and 52% of that obtained from the reference thermal cycler at ligation times of 2 min and 15 s, respectively The lower net LDR yield at higher flow rates results from thermal nonequilibrium conditions of the fluidic packet traversing through the isothermal zones with the set temperature. Finite element simulations in our previous studies[48] indicate that the reaction mixture reaches thermal equilibrium with the set temperature at a flow rate of 0.22 μL/min, which provides an effective residence time of ~2 min within the nominal ligation zone. A flow rate of 1.8 μL/min corresponds to a 15-s effective residence time and requires at least 1 s to reach the set temperature. Although the on-chip LDR produced slightly lower product yields using short ligation times, the processing time was shorter than that of the reference cycler. The total processing time for 20 thermal cycles was 6.5 min using the CFLDR microchip (3 6-s denaturation, 15.8 s for renaturation, ligation, and heat transition; cycling rate ~19 5 s/cycle) while ~25 min was required using the reference cycler (30-s denaturation, 15 s for renaturation and ligation, and ~29 s for heat transition, cycling rate of ~74 s/cycle), when the same nominal ligation time of 15 s was selected for both platforms

(64) Cheng, J. Schöffner, M. A., Hvichia, G. E., Kricka, L. J. Wilding, P. Nucleic Acids Res. 1996 24, 380–385

On-Chip Multiplexed LDR. The sensitivity of PCR/LDR was determined by reconstructing samples containing known amounts of mutant DNA (derived from cell lines) with a fixed copy number of wild-type DNA. Samples (mutant G12D, which is the most commonly found mutation in K-ras, and wild-type G12) were PCR amplified independently and then mixed in known copy numbers with the LDR carried out in the continuous-flow microchip allowing a range of mutant-to-wild-type ratios to be evaluated We examined signals generated from correct ligation of the mutant template and backgrounds generated from misligations occurring from the wild-type template using a multiplex LDR primer set (discriminating primers to detect G12D, G12A, and G12V and DY782-labeled common primers; see the left scheme of Figure 4A). Panels A and B in Figure 4 represent gel images showing LDR products and an electropherogram of these multiplexed LDRs. No product was seen when the sample lacked any template in the reaction (see lane 6). When 10 nM wild-type template (G12) was added into the reaction mixture, a signal indicating C·A, C·C, and C·T mismatched ligation products was observed (see lane 5). The gel band produced from the mutant allele (matched ligation product) was distinguished from the noise (mismatched ligation product) at a SNR > 2 with a mutant to wild-type ratio of 1:100 (compare lanes 2 and 5). The high fidelity of LDR in distinguishing mutant target (complete match between discriminating primer and target) in the presence of a high molar excess of wild-type sequences is due to the ability of the thermostable ligase to rapidly dissociate from substrates containing mismatches. Even if a misligation event does occur at an early LDR cycle, the product does not undergo subsequent amplification as is the case in PCR. Therefore, in contrast to allele-specific PCR, PCR/LDR does not selectively amplify low-level polymerase errors and, hence, reduces the chance of false positives. Figure 4C shows the difference in product yield and ligation fidelity between Tth ligase and Taq ligase. Taq ligase showed a slightly larger product yield compared to Tth ligase with no significant difference in ligation fidelity

Hybridization Stringency Using LDR Buffer. SSPE or sodium chloride–sodium citrate (SSC) buffer along with SDS has been conventionally used as a hybridization buffer for Southern and Northern blots in order to increase hybridization stringency. The use of such hybridization buffers following LDR would require incorporation of additional microchannels, microfluidic mixers, microvalves, or perhaps DNA extraction chambers prior to hybridization since the composition of the LDR buffer (see Experimental Section) is different from that of the hybridization buffer, which may adversely affect hybridization stringency. We investigated the effects of the buffer composition on the hybridization (see Figure 5) The compatibility of the LDR buffer with the low-density zip code universal array was evaluated to determine whether post-LDR mixing with the appropriate hybridization buffer would be necessary A LDR was performed using the reference cycler and discriminating primer to detect wild-type template G12 and a common primer, which should produce a matched LDR product that has a complementary sequence to zip code probe 1. After the reaction, the product solution was diluted with 5× SSPE–0.1% SDS buffer at a volumetric ratio of 50:50 (see Figure 5A). On the other hand, the same product solution was diluted with the LDR cocktail at a volumetric ratio of 50:50 (see Figure 5B) to produce the same concentration of solution complementary
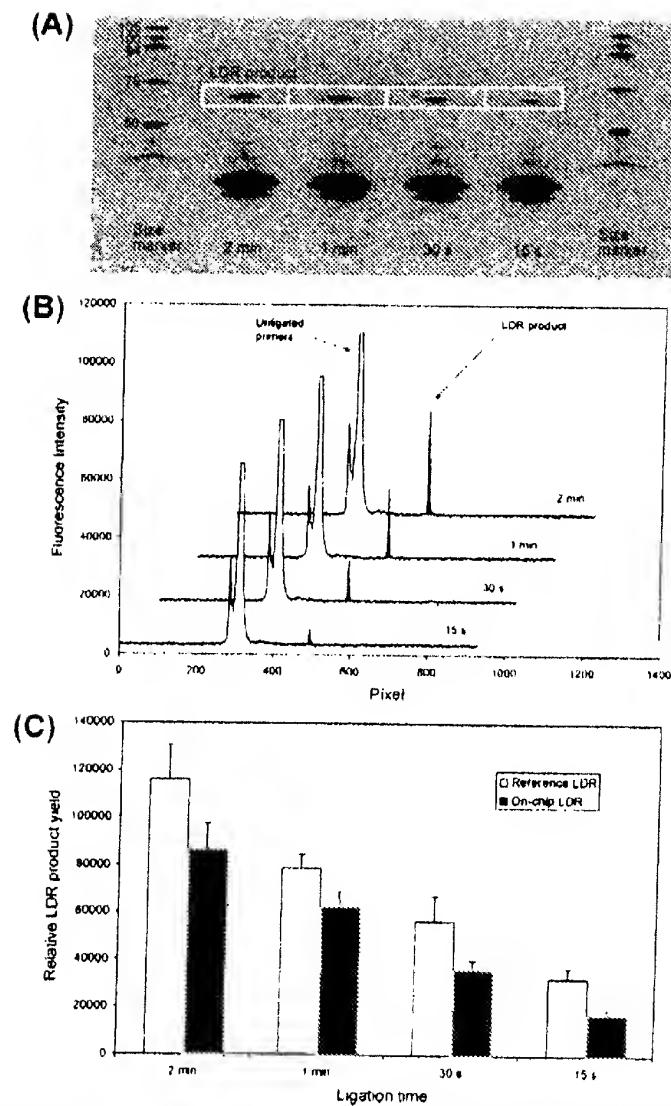
Figure 3. Effects of ligation time on LDR product yield (A) Fluorescence images of the on-chip LDR products sorted electrophoretically using a 5.5% polyacrylamide gel matrix (B) Fluorescence was integrated over the indicated area in (A) with image integration software (C) Comparison of the LDR product yield between the continuous-flow LDR chip and the reference thermal cycler The ligation times for the LDR chip were estimated from the linear velocities of the fluids

sequences The solutions were hydrostatically moved through the microarray chip at a volumetric flow rate of 1.8 μL/min for 5 min set to 55 °C. After hybridization, the chip was rinsed with wash buffer and imaged using the near-IR scanner. No significant differences in hybridization stringency were observed between

the LDR buffer and the hybridization buffer The need for high stringency in the hybridization is relaxed using the universal array format employed herein, since there is large sequence variability between different elements (zip code probes) of the array and the mutation screening assay is decoupled from hybridization.
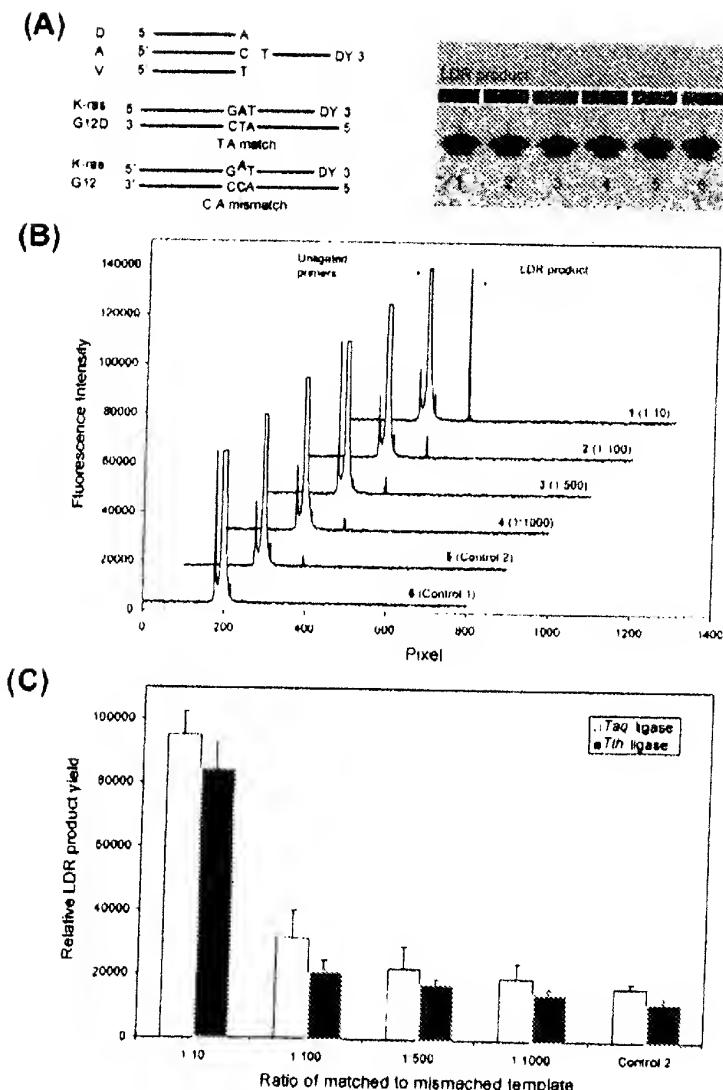
Figure 4. On-chip LDRs of K-*ras* G12D mutant and G12 wild-type DNAs with different ratios of mutant to wild-type sequences. (A) Left, schematic diagram of mutant G12D and normal G12 double-stranded templates with the three discriminating (D, A, end V) and one common primer used in the LDR; right, fluorescence images of the on-chip LDR products enalyzed electrophoretically using a 5 5% polyacrylamide gel matrix. Lene 1, 1 nM mutant in 10 nM wild-type template (1 10); lane 2, 0 1 nM mutant in 10 nM wild-type template (1 100); lanes 3, 0 02 nM mutant in 10 nM wild-type template (1 500) lene 4, 0 01 nM mutant in 10 nM wild-type template (1 1000); lane 5, 10 nM wild-type DNA only (negative control 2). lane 6. no template DNA (negative control 1) (B) Fluorescence was integrated over the indicated areas in (A) with image integration software (C) Comparison in the relative LDR product yield and ligation fidelity between *Taq* end *Tth* ligase enzymes

---

Detecting K-*ras* Mutations Using the PC/PDMS/PMMA Microfluidic Assembly. As discussed previously, the mutation signal from matched product was distinguished from noise (signal

from mismatched product) at a sensitivity level of 1:100 using gel electrophoresis. However, the allelic composition of the specific point mutation could not be identified when unknown templates
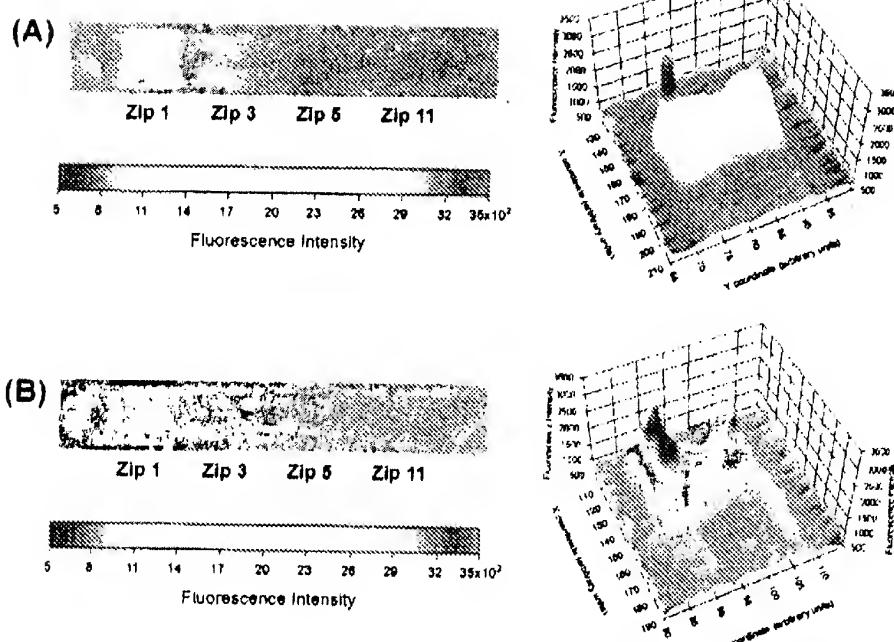
Figure 5. Influence of the LDR cocktail on hybridization stringency. The LDR was carried out in a conventional block thermal cycler using the G12 wild-type DNA only. The LDR generates a product that hybridizes to zip code probe 1. The solutions subjected to hybridization analysis were (A) LDR product/hybridization buffer (5× SSPE–0.1% SDS) at a volume ratio of 50:50 or (B) LDR product/LDR mixture, 50:50 (v/v). Shown on the left is a two-dimensional image of the entire hybridization channel and on the right is a three-dimensional representation of the zip code probe 1 fluorescence image. The solutions were flowed through the chip using a syringe pump at a volumetric flow rate of 1.8 µL/min. After the hybridization, the chip was rinsed with a wash buffer and imaged using the near-IR scanner

are examined using gel electrophoresis because all of the discriminating primers have the same size (47-mer). Providing differences in primer length for the discriminating primers could make the identification possible,[21,24] although gel electrophoresis is a time-consuming step and is limited in the number of mutations and their allelic composition that can be simultaneously analyzed. Alternatively, the identification of successful LDRs can be carried out using a microarray by displaying positions of fluorescence signatures generated on a solid support where known probes (zip codes) are tethered.

As demonstrated in Figure 3, the larger the flow rate of sample through the CFLDR device, the smaller the amount of product generated due to reduced ligation time. However, larger flow rates can accelerate hybridization kinetics reducing analysis time.[24] Therefore it was necessary to balance processing speed with detection sensitivity, both of which depend on the volume flow rate at which the biochip was operated. The hybridization assay was performed at a flow rate of 1.8 µL/min using the biochip in order to detect the presence of low-abundant G12D mutations in the presence of a majority of wild-type G12 sequences (see Figure 6). This volume flow rate (ligation time, 15 s) was chosen since it provided adequate LDR product for analysis (see Figure 3).

When only wild-type templates G12 were added into the reaction cocktail with one discriminating primer, which was specific for the G12 sequence, and one common primer, only C·G matched products were produced as seen by the corresponding signal generated at zip code 1 (1, see plot 1 in Figure 6B). When all four discriminating primers, which can detect alleles G12, G12D, G12A, and G12V, were used with G12, small amounts of mismatched products (C:A, C:C, and C:T mismatches) were produced as well as a vast majority of C:G matched products. The mismatched products hybridized to their complementary zip code probes (2, see graph 2 in Figure 6B), providing background noise when attempting to detect target mutations. When 1 nM K-*ras* G12D mutant was added into the reaction mixture along with 10 nM G12, large amounts of T·A matched LDR products were generated and hybridized to the appropriate zip code probes (zip code 3) (3, see graph 3 in Figure 6B) The signal intensity was well above background produced from C:A mismatched products (compare graph 3 to graph 2 for zip code probe 3 in Figure 6B). The higher signals at zip code probes 5 and 11 were produced by generation of T:C and T:T mismatched products due to the presence of the mutant G12D allele (compare graph 3 to graph 2 at zip code probe 5 and zip code probe 11). Using our image
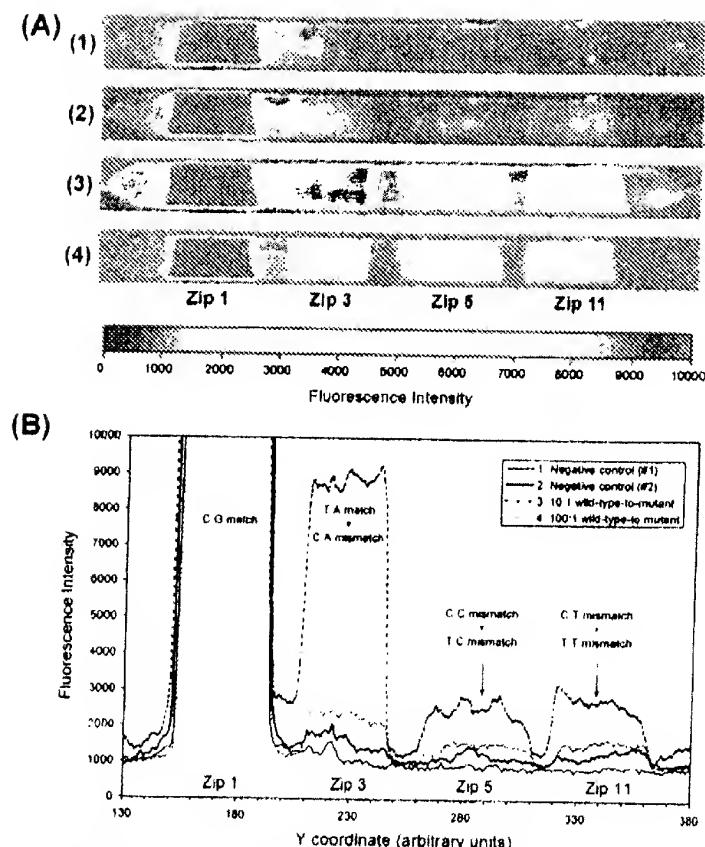
Figure 6. Mutational analysis of point mutations in K-*ras* using different wild-type-to-mutant ratios. The LDRs were analyzed using the LDR/zip code array biochip assembly (A) 2D image obtained with the near-IR scanner and (B) line scans through the microchamber image 1, negative control using one discriminating primer to detect G12 and one common primer (control 1), 2, negative control using the set of four discriminating primers to detect G12, G12D, G12A, and G12V and one common primer (control 2), 3, positive control using the set of four discriminating primers and one common primer with a template ratio of 10:1 (wild-type G12-to-K-*ras* G12D), 4, positive control using the set of four discriminating primers and one common primer with a template ratio of 100:1 (wild-type G12-to-K-*ras* G12D). The LDR cocktail was pumped through the assembly using a flow rate of 1.8 μL/min (55 °C hybridization temperature).

analysis software, we integrated the total fluorescence counts at zip code 3 (mutation signal, T A match) and zip code 5 (background generated from misligation, C C + T C mismatch) at a level of 1 100 mutation to wild-type sequences (4, see graph 4 in Figure 6B). This analysis provided a count number of 520 333 for zip code 3 (corrected for autofluorescence from PMMA substrate at locations of the array not bearing zip code probes) and 211 823 counts for zip code 5 (corrected for PMMA autofluorescence). These numbers provided a signal-to-background ratio of ~2.5 at the mutant to wild-type ratio of 1:100. Implementing longer hybridization times (>5 min, i.e., lower volume flow rates) did not seem to enhance the relative detection sensitivity of these mutations because the background due to the mismatched products increased at the same level as the matched ligation

products. The time required for scanning the entire hybridization microchamber (500 μm wide × 6.7 mm long) was <10.3 min with a step resolution of 25.4 μm. A 50.8-μm step resolution, which can complete the scanning in 2.6 min, also provided well-resolved images because of the low-density array format used herein.

CONCLUSIONS

We have fabricated a flow-through biochip that consisted of two different materials, PC and PMMA, for the detection of low-abundant DNA mutations in gene fragments (K-*ras*) that carry point mutations with high diagnostic value for colorectal cancers. The microchips possessed discrete functions, i.e., the PC chip for continuous-flow LDR and the PMMA chip for universal zip code array readout. The physiochemical properties of these

materials (high glass transition temperature for PC and lower fluorescence background and minimal nonspecific DNA adsorption for PMMA) matched the operational needs for each chip. The two microchips were assembled in a three-dimensional architecture using an interconnect, which was fabricated by laser drilling microchannels and included a microgasket fabricated from an elastomer (PDMS). The integrated biochip was micromanufactured using several different techniques, such as hot embossing for producing the intrachip fluidic network and laser ablation for producing the interchip fluidics. Our experiments indicated the ability to detect one mutant sequence in ~100 normal sequences using this integrated device. The miniaturized reaction channel and the continuous-flow operation of the LDR microthermal cycler accelerated the reaction primarily through its enhanced thermal management capabilities. The zip code array constructed in the microfluidic channel displayed improved hybridization kinetics compared to conventional nonflow formats due to enhanced mass transport to its surface and minimal diffusional distances. Because of these attributes, the assay could be carried out rapidly: 6.5 min for on-chip LDR, 10 min for washing, and 2.6 min for scanning (~19.1 min total). This is a significant reduction in processing time when compared to previous work where all of these steps were carried out using conventional instrument platforms: 95 min for LDR, 15-min preincubation, 120 min for hybridization, 10 min for washing, and 30 min for imaging (total processing time ~270 min).[15] In addition to enhanced processing speed, the reagent volume required for the microchip format was reduced >10-fold compared to the conventional format. Our biochip could be manufactured at relatively low cost (~$0.33/chip, materials cost only) due to the use of replication technology and polymer parts and contained only passive elements, a particularly attractive format for clinical applications where disposable-type devices are a necessity to eliminate false positives arising from carryover effects. We are currently working on the incorporation of a primary PCR amplification step into the stacked fluidic network as well as cell lysis and DNA extraction to provide a rapid and high-sensitivity PCR/LDR/hybridization biochip to provide multiplexed processing for massive parallel screening of single-nucleotide changes in cancer cells for clinical staging of solid tumors.

Exhibit 8:   Favis et al., "Universal DNA Array Detection of Small Insertions and Deletion in BRCA1 and BRCA2," *Nat. Biotech.* 18:561–564 (2000)

Universal DNA array detection of small insertions and deletions in *BRCA1* and *BRCA2*

Reyna Favis[1], Joseph P. Day[1], Norman P. Gerry[1], Catherine Phelan[2], Steven Narod[2], and Francis Barany[1]*

[1]Department of Microbiology, Hearst Microbiology Research Center and Strang Cancer Prevention Center, Joan and Sanford I. Weill Medical College of Cornell University, 1300 York Ave., Box 62, New York, NY 10021, USA. [2]Center for Research on Women's Health, University of Toronto, Toronto, M5G 1N8, Canada. *Corresponding author (barany@mail.med.cornell.edu).

Array-based mutation detection methodology typically relies on direct hybridization of the fluorescently labeled query sequence to surface-bound oligonucleotide probes. These probes contain either small sequence variations or perfect-match sequence. The intensity of fluorescence bound to each oligonucleotide probe is intended to reveal which sequence is perfectly complementary to the query sequence[1]. However, these approaches have not always been successful, especially for detection of small frameshift mutations. Here we describe a multiplex assay to detect small insertions and deletions by using a modified PCR to evenly amplify each amplicon (PCR/PCR)[2], followed by ligase detection reaction (LDR)[3]. Mutations were identified by screening reaction products with a universal DNA microarray[4], which uncouples mutation detection from array hybridization and provides for high sensitivity. Using the three *BRCA1* and *BRCA2* founder mutations in the Ashkenazi Jewish population (*BRCA1* 185delAG; *BRCA1* 5382insC; *BRCA2* 6174delT)[5] as a model system, the assay readily detected these mutations in multiplexed reactions. Our results demonstrate that universal microarray analysis of PCR/PCR/LDR[3] products permits rapid identification of small insertion and deletion mutations in the context of both clinical diagnosis and population studies.

We developed this method as an alternative to direct probe/target hybridization, which when used to study mutations in exon 11 of BRCA1[3], failed to detect the presence of the 1128insA mutation. Insertion and deletion mutations in *p53* also proved refractory to detection by direct hybridization: none of five frameshift mutations in exons 2–11 of *p53* were detected[6]. These inaccuracies may result from disruption of secondary structure: perfect match sequence may assume a secondary structure that is eliminated in variant sequences. This structural change may lead to binding of variant target to perfect-match probe with higher binding affinity than true perfect-match target[6]. Another consequence of direct hybridization is the formation of stable duplexes by looping out of noncomplementary sequence during hybridization. Either of these illegitimate hybridizations could produce false negative signals on an array.

Direct hybridization of mutation-containing target sequence to sequences on the array has an additional shortcoming: the difficulty of simultaneously assaying sequence tracts with localized regions of high G•C and A•T content[6]. This may also lead to false negative signals. An alternative array detection scheme based on single-nucleotide extension also fails to detect slippage of mononucleotide repeat sequences[7]. To overcome these deficits, we have developed a universal microarray wherein signal detection is completely uncoupled from mutation identification.

In screening more than 80 samples, we successfully detected the small insertions and deletions found in the three Ashkenazi *BRCA1* and *BRCA2* mutations. The exons in which these mutations are located were selectively amplified from genomic DNA samples using a multiplex PCR reaction (PCR/PCR/LDR) designed to minimize primer-specific differences in amplification efficiency, as outlined in Figure 1. Specific insertions and deletions were first detected by multiplex LDR and subsequently separated by electrophoresis. A representative gel is shown in Figure 2A. Each lane contains comparable amounts of each LDR product, indicating equivalent amplification efficiency for each sample. All wild-type signals are blue in color (FAM label), whereas the mutant signals are green (TET label). This arrangement enables rapid visual screening of samples for the presence or absence of mutations.

Ligase detection reaction had been shown to be a sensitive assay for detecting point mutations[3]. To determine the limits of sensitivity of this assay for frameshift mutations, we performed a simulation experiment wherein a defined quantity of genomic DNA with a known mutation was diluted from 1:2 to 1:20 with wild-type genomic DNA. The mixed DNA samples were amplified as before and then subjected to multiplex LDR to detect the presence of the
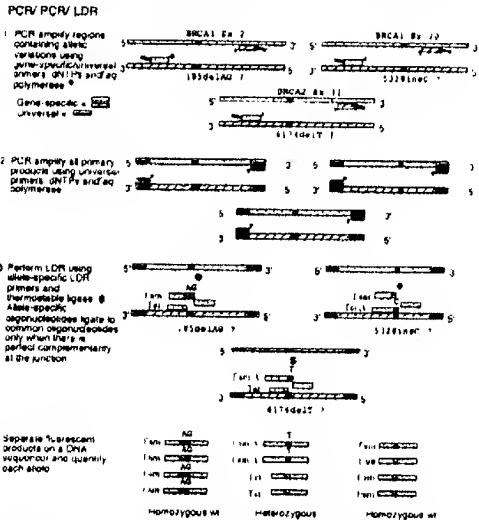
PCR/ PCR/ LDR

Figure 1. Outline of the PCR/PCR/LDR method for detection of mutations in *BRCA1* and *BRCA2*. Multiplex amplification of the relevant exons is carried out to ensure equal amplification of all products. A limited number of PCR cycles is performed using gene-specific primers, with further rounds of amplification primed from the universal sequences located at the extreme 5' ends of the PCR primers. LDR is then used to detect both wild-type and mutant versions of the queried sequence. The ligation oligonucleotides hybridize to both wild-type and mutant products, but ligate only when both primers are perfectly matched with no gaps or overlaps. Products are electrophoretically separated or hybridized to a microarray for identification.

TECHNICAL REPORTS

A

B
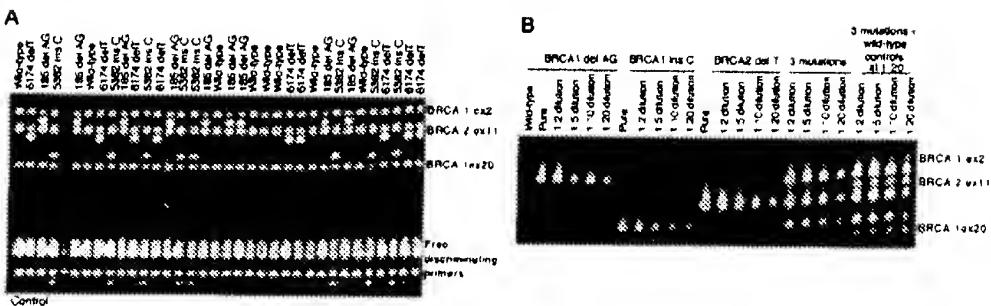
Figure 2. Multiplex LDR detection of three specific mutations in *BRCA1* and *BRCA2* in a gel-based assay. (A) Representative LDR gel for detection of the three *BRCA* mutations. By fluorescently end-labeling the discriminating upstream oligonucleotides with either FAM (for wild-type) or TET (for mutant), and by adding "tails" of different lengths to LDR oligonucleotides, ligation products can be distinguished based on label end migration on a polyacrylamide gel. Wild-type products are identified at the right side of the gel. Mutant products are identified at the top of the gel. (B) *BRCA1* and *BRCA2* mutation detection on pooled samples of DNA. DNA samples with known mutations were diluted into wild-type DNA before amplification. Ligation products from multiplex LDR are shown for each dilution. *BRCA1* del AG, *BRCA1* ins C, and *BRCA2* del T = only one mutation present; multiplex LDR directed against only mutant sequences using 500 fmol of each LDR oligonucleotide. 3 mutations = all three mutations present; multiplex LDR directed against mutant sequences only using 500 fmol of each LDR oligonucleotide. 3 mutations + wild-type controls at 1:20 = all three mutations present; multiplex LDR directed against both mutant and wild-type sequences using 500 fmol and 25 fmol of each LDR oligonucleotide, respectively.

mutation under both single- and multiple-mutation conditions Figure 2B shows that each of the three *BRCA1* and *BRCA2* mutations can be detected when diluted 1:20, even in a multiplex format also containing diluted wild-type LDR oligonucleotide controls and all three mutations

The ability of PCR/PCR/LDR to rapidly process samples in quantity was tested in a blind study using 249 DNA samples derived from Ashkenazi Jewish individuals. Although the concentration of the DNA ranged from 15 ng μl⁻¹ to 3,000 ng μl⁻¹, samples were assembled in 9 × 9 configurations without prejudice and pooled across rows and down columns. After PCR amplification, these pooled samples were then subjected to LDR as described above. We confirmed mutations by analyzing individual DNA samples residing at the intersection of columns and rows found to be positive for mutation. This strategy allowed unique typing of each sample. Of the 249 samples, 248 were correctly identified. The single sample that was incorrectly identified as wild type proved to be too dilute and fell below the limits of detection when mixed with nine other samples of higher concentration. The number of individual reactions carried out was reduced from 249 to 96 by this strategy (55 pooled samples and 41 individual samples used for confirmation).

Since gel electrophoresis limits the number of mutations that can be simultaneously analyzed, analysis of the LDR products was transferred to a zip-code-based microarray system. Zip-code sequences consist of 24 bases that are assembled from a set of 36 tetramers. Each tetramer differs from the others by at least two bases and is neither self-complementary, nor complementary to any other tetramer. Similarly, each zip code differs from the others by at least three tetramer units. The end result of this design are sequences that have comparable behavior in terms of the thermodynamics and kinetics of hybridization, while simultaneously maintaining distinct chemical identities that prevent cross-hybridization. Each zip code becomes associated with a particular mutation when its complement is attached to the nonreactive 3′ end of a common LDR oligonucleotide that is specific for a particular mutation. As LDR is a specific reaction[10,11], ligation products fusing the fluorescent- and zip-code complement-containing oligonucleotides are produced only in the presence of a template sequence that is an exact match at the ligation junction Since the zip codes reside at defined addresses on the microarray surface, each zip-code complement will direct specific wild-type or mutant LDR products to a unique address on the array.

Hence, zip-code hybridization on an array allows complete physical separation of products and obviates the need for gel electrophoresis

Preliminary microarray experiments using primers designed in the gel-based format (i.e., differentially labeled discriminating primers and identical common primers) demonstrated that wild-type and mutant versions of the three *BRCA* sequences were readily
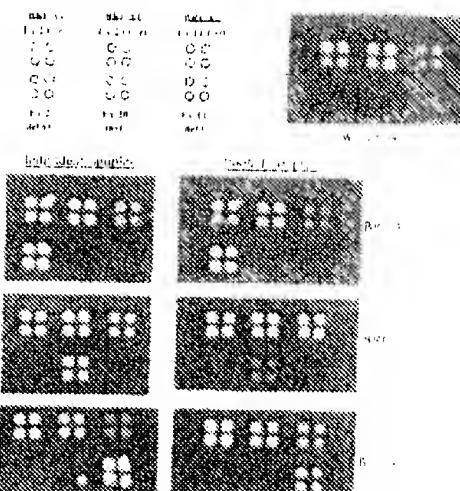
Figure 3. LDR detection of three specific mutations in *BRCA1* and *BRCA2* on an addressable universal microarray. The diagram at upper left shows the assignment of each control and mutant sequence to specific addresses on the array surface. Control signals are directed to the upper three addresses; mutant signals are assigned to the lower three. The upper right image shows signal produced by a wild-type DNA. Left panel: representative hybridizations for individual DNA samples. Right panel: representative hybridizations for each mutation using pooled samples of DNA from Ashkenazi individuals. The mutations are identified on the extreme right.

Table 1. PCR and LDR oligonucleotides*

Oligonucleotide description	Sequence
PCR primers	
Universal primer A (Uni A)	5'-ggagcacgctatcccgttagac-3'
Universal primer B (Uni B)	5'-cgctgccaactaccgcccelg-3'
BRCA1 exon 2 forward	5'-Uni A-TCATTGGAACAGAAAGAAATGGATTTATC-3'
BRCA1 exon 2 reverse	5'-Uni B-TCTTCCCTAGTATGTAAGGTCAATTCTGTTC-3'
BRCA1 exon 20 forward	5'-Uni A-ACTTCCATTGAAGGAAGCTTCTCTTTC-3'
BRCA1 exon 20 reverse	5'-Uni B-ATCTCTGCAAAGGGGAGTGGAATAC-3'
BRCA2 exon 11 forward	5'-Uni A-CAAAATATGTCTGGATTGGAGAAAGTTTC-3'
BRCA2 exon 11 reverse	5'-Uni B-TTGGAAAAGACTTGCTTGGTACTATCTTC-3'

LDR gel-based assay

Discriminating oligonucleotides:	
BRCA1 ex 2 wild-type position 185	5'-Fem-aaCATTAATGCTATGCAGAAAATCTTAGAG-3'
BRCA1 ex 2 position 185 del AG mutation	5'-Tet-GTCATTAATGCTATGCAGAAAATCTTAG-3'
BRCA1 ex 20 wild-type position 5382	5'-Fam-CCAAAGCGAGCAAGAGAATCC-3'
BRCA1 ex 20 position 5382 ins C mutation	5'-Tet-aaCAAAGCGAGCAAGAGAATCCC-3'
BRCA2 ex 11 wild-type position 8174	5'-Fem-caCTTGTGGGATTTTTAGCACAGCAAGT-3'
BRCA2 ex 11 position 6174 del T mutation	5'-Tet-TACTTGTGGGATTTTTAGCACAGCAAG-3'

LDR Common oligonucleotides:	
BRCA1 ex 2 position 185	5'-P-TGTCCCATCTGGTAAGTCAGCACAAAC-B-3'
BRCA1 ex 20 position 5382	5'-P-CAGGACAGAAAGGTAAAGCTCCCTC-B-3'
BRCA2 ex 11 position 6174	5'-P-GGAAAATCTGTCCAGGTATCAGAT-B-3'

LDR microarray assay

Discriminating oligonucleotides:	
BRCA1 ex 2 control	5'-Cy3-TGCATAGGAGATAATCATAGGAATCC-3'
BRCA1 ex 2 position 185 del AG mutation	5'-Cy3-GTCATTAATGCTATGCAGAAAATCTTAG-3'
BRCA1 ex 20 control	5'-Cy3-CCTCTGACTTCAAAATCATGCTG-3'
BRCA1 ex 20 position 5382 ins C mutation	5'-Cy3-CAAAGCGAGCAAGAGAATCCC-3'
BRCA2 ex 11 control	5'-Cy3-CTTCCCTATACTACATTTACATATATCTGAAG-3'
BRCA2 ex 11 position 8174 del T mutation	5'-Cy3-TACTTGTGGGATTTTTAGCACAGCAAG-3'

Common oligonucleotides for controls:	
BRCA1 exon 2 control + cZip 1	5'-P-CAAATTAATACACTCTTGTGCTGACTTACCA-cgc agatttgcgctggatticaa-B-3'
BRCA1 exon 20 control + cZip 2	5'-P-AAAGAAACCAAACACAACCCCATCAG-ttcgccglc glgtaggctttcaa-B-3'
BRCA2 exon 11 control + cZip 3	5'-P-TTTCCAAACTAACATCACAAGGTGATATTT-ccglaa gcccgtatggcagetcaa-B-3'

Common oligonucleotides for mutations:	
BRCA1 exon 2 position 185 + cZip 9	5'-P-TGTCCCATCTGGTAAGTCAGCACAAAC-catcgtcc ctttcgatgggatcaa-B-3'
BRCA1 exon 20 position 5382 + cZip 10	5'-P-CAGGACAGAAAGGTAAAGCTCCCTC-caaggcacg tcccagacgcatcaa-B-3'
BRCA2 exon 11 position 6174 + cZip 1 1	5'-P-GGAAAATCTGTCCAGGTATCAGAT-gcacgggagctg acgacgtgtcaa-B-3'

*LDR oligonucleotides are depicted in the 5' to 3' orientation. In all cases, upper-case bases indicate genomic sequence, lower-case bases indicate nongenomic sequence (either complementary zip codes or universal primers); bold lower-case bases indicate nongenomic bases added to control the size of the final product, P, phosphate; B, blocking group. cZip, complementary zip code.

detected on the array (data not shown). In this version, both types of sequences were directed to the same addresses (e.g., BRCA1 185delAG and BRCA1 185 wild type were both directed to zip code 1). Although this format proved successful, PCR/PCR/LDR has the potential of detecting hundreds of mutations in a single-tube reaction, and this design does not make optimal use of the array for such large-scale mutation detection experiments. In order to establish the experimental paradigm for future studies, we expanded the addressable format by choosing a sequence in each of the amplicons to use as a control LDR product. Thus, rather than require detection of wild-type sequences for each mutant LDR product, this format uses a single product to serve as a positive control for multiple different sequence variants within an amplicon. One advantage of this format

is that it minimizes oligonucleotide synthesis; additionally, the use of each of the 64 positions is maximized. Since the number of LDR products that can be detected at a single address is limited by the number of currently available spectrally separated fluorescent labels, confining the control to a specified region of the array permits one more sequence variant to be detected at each remaining address. In the experiments described here, control and mutant LDR products for a queried position were directed to six separate addresses on a 64-position array. As a control for reproducibility, each address was spotted in quadruplicate.

All three frameshift mutations were detectable by hybridization to the universal array (Fig. 3). Only combinations of the six possible addresses were visible following hybridization, and no additional signals were detected at any of the unused addresses. Thus, zip-code hybridization is very specific. Control and mutant signals were clearly present for each of the mutations derived from samples of DNA from single individuals (Fig. 3, left panel). Pooled DNA used in analyzing the 249 DNA samples described above produced signals for mutations identical to those found in the gel-based assay (Fig. 3, right panel). In each case, the array reproduced the result of the gel.

In addition to allowing rapid screening of multiple individuals, PCR/LDR can also identify individuals with multiple mutations. A previous blind study performed using 144 tumor samples successfully detected all mutations out of 19 possible closely clustered single-base mutations that occur at codons 12, 13, and 61 of the K-ras gene[3]. In addition, PCR/LDR successfully detected length polymorphisms resulting from small insertions and deletions in both dinucleotide repeats and mononucleotide repeats in the APC (APC1307K) and the TGF-β type II receptor genes[2,11]. Results from these studies were corroborated by and exceeded detection by direct sequencing.

Recently, PCR/LDR was used in combination with the universal DNA array to detect K-ras mutations in tumor and cell line DNA[4]. Together, these approaches can provide the high sensitivity required to detect low-level mutations (unavailable in other arrays) and the speed required to rapidly assay large numbers of clinical samples. Indeed, our experiment analyzing mutations in pooled DNA samples emphasizes the utility of this approach in screening large populations. The ability to detect mutations in pooled samples will facilitate large-scale correlative studies, where unclassified polymorphisms are compared in diseased and healthy cohorts to determine if particular polymorphisms contribute to development of disease[4]. Finally, since multiple loci can be simultaneously assayed, it may be possible to investigate genetic interactions by analyzing combinations of alleles in both diseased and unaffected cohorts.

TECHNICAL REPORTS

Experimental protocol

PCR/PCR amplification. Polymerase chain reaction was carried out as a single-tube, multiplex reaction to simultaneously amplify *BRCA1* exons 2 and 20 and *BRCA2* exon 11. Genomic DNA was extracted from blood samples of Ashkenazi Jewish individuals and amplified in a 25 μl reaction mixture containing 100 ng of DNA, 400 μM of each dNTP, 1× PCR buffer II (10 mM Tris-HCl pH 8.3 at 25°C, 50 mM KCl) supplemented with 4 mM MgCl₂, 1 U AmpliTaq Gold, and 2 pmol of each gene-specific primer bearing either universal primer A or B on the 5' ends (see Table 1 for sequences). The reaction was overlaid with mineral oil and preincubated for 10 min at 95°C. Amplification was performed for 15 cycles as follows: 94°C for 15 s, 65 °C for 1 min. A second 25 μl aliquot of the reaction mixture was added through the mineral oil containing 25 pmol each of universal primers A and B. Cycling was repeated using 55°C annealing temperature. The reaction was next digested with a 2 μl solution of 1 mg ml⁻¹ proteinase K–50 mM EDTA at 55°C for 10 min. Proteinase K was eliminated by a final incubation at 90°C for 15 min.

To test mutation detection in pooled diluted samples, simulation experiments were performed. DNA samples with known mutations were diluted 1:2, 1:5, 1:10, and 1:20 with wild-type DNA before PCR amplification, and then subjected to LDR. For pooling blinded Ashkenazi Jewish DNA samples, the tubes containing the DNAs were assembled into a 9 × 9 gridded format and aliquots from each tube were combined across the rows and then down the columns to produce one tube of combined DNA for each row and each column. The pooled DNA was then subjected to amplification and LDR. See Table 1 for the PCR and LDR primer sequences.

LDR conditions. Oligonucleotide synthesis and purification were carried out as described⁴, using a Pixsys5500 robot enclosed in a humidity chamber (Cartesian Technologies, Irvine, CA). Briefly, LDR reactions were hybridized in 32 μl containing 300 mM 2-[N-morpholino] ethanesulfonic acid, pH 6.0, 10 mM MgCl₂, 0.1% SDS at 65°C for 1 h in a rotating chamber. After washing in 300 mM bicine, pH 8.0, 10 mM MgCl₂, 0.1% SDS for 10 min at 25°C, the array

was imaged on an Olympus Provis AX70 microscope using a 100 W mercury burner, a Texas Red filter cube, and a Princeton Instruments TEK512/CCD camera. The 16-bit gray-scale images were captured using MetaMorph Imaging System (Universal Imaging, West Chester, PA) and rescaled to more narrowly bracket the LDR signal before conversion to 8-bit gray-scale. The 8-bit images were colored using Adobe Photoshop to render the Cy3 signal red.

1 Kozal, M.J. et al. Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nat. Med.* 2, 753–759 (1996).

2 Belgreder, P., Marino, M., Lubin, M. & Barany, F. A multiplex PCR-ligase detection reaction assay for human identity testing. *Genome Sci. Technol.* 1, 77–87 (1996).

3 Khanna, M. et al. Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumors. *Oncogene* 18, 27–38 (1999).

4 Gerry, N. et al. Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* 292, 251–262 (1999).

5 Rahman, N. & Stratton, M. The genetics of breast cancer susceptibility. *Annu. Rev. Genet.* 32, 95–121 (1998).

7 Hacia, J., Brody, L., Chee, M., Fodor, S. & Collins, F. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat. Genet.* 14, 441–447 (1996).

8 Ahrendt, S. et al. Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array. *Proc. Natl. Acad. Sci. USA.* 96, 7382–7387 (1999).

6 Hacia, J. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat. Genet.* (Suppl.) 21, 42–47 (1999).

9 Syvanen, A.C. & Landegren, U. Detection of point mutations by solid-phase methods. *Hum. Mutat.* 3, 172–179 (1994).

10 Tong, J., Cao, W. & Barany, F. Biochemical properties of a high fidelity DNA ligase from *Thermus* species AK16D. *Nucleic Acids Res.* 27, 788–794 (1999).

11 Luo, J., Bergstrom, D.E. & Barany, F. Improving the fidelity of *Thermus thermophilus* DNA ligase. *Nucleic Acids Res.* 24, 3071–3078 (1996).

12 Zirvi, M. et al. Ligase-based detection of mononucleotide repeat sequences. *Nucleic Acids Res.* 27, 640 (1999).

13 Zirvi, M., Bergstrom, D.E., Saurage, A.S., Hammer, R.P. & Barany, F. Improved fidelity of thermostable ligases for detection of microsatellite repeat sequences using nucleoside analogs. *Nucleic Acids Res.* 27, e41 (1999).

14 Chakravarti, A. Population genetics – making sense out of sequence. *Nat. Genet.* 21, 56–60 (1999).

15 Luo, J. & Barany, F. Identification of essential residues in *Thermus thermophilus* DNA ligase. *Nucleic Acids Res.* 24, 3079–3085 (1996).

16 Barany, F. & Gelfand, D. Cloning, overexpression, and nucleotide sequence of a thermostable DNA ligase gene. *Gene* 109, 1–11 (1991).

Exhibit 9:   Fouquet et al., "Rapid and Sensitive p53 Alteration Analysis in Biopsies from Lung Cancer Patients Using a Functional Assay and A Universal Oligonucleotide Array: A Prospective Study," *Clin Cancer Res* 10:3479–3489 (2004)

Rapid and Sensitive p53 Alteration Analysis in Biopsies from Lung Cancer Patients Using a Functional Assay and A Universal Oligonucleotide Array: A Prospective Study

Coralle Fouquet,[1] Martine Antoine,[1,2]
Pascaline Tisserand,[1] Reyna Favis,[4]
Marie Wislez,[1,3] Fréderic Commo,[2]
Nathalie Rabbe,[1,3] Marie France Carette,[1,3]
Bernard Milleron,[1,3] Francis Barany,[4]
Jacques Cadranel,[1,3] Gérard Zalcman,[1] and
Thierry Soussi[1]

[1]Laboratoire de génotoxicologie des tumeurs, Paris, France, [2]Service d'Anatomie Pathologique, and [3]Service de Pneumologie et de Radiologie, Hôpital Tenon, Paris, France, and [4]Department of Microbiology, Cornell University, New York, New York

ABSTRACT

Purpose: Molecular profiling of alterations associated with lung cancer holds the promise to define clinical parameters such as response to treatment or survival. Because <5% of small cell lung cancers and <30% of non-small cell lung cancers are surgically resectable, molecular analysis will perforce rely on routinely available clinical samples such as biopsies. Identifying tumor mutations in such samples will require a sensitive and robust technology to overcome signal from excess amounts of normal DNA.

Experimental Design: p53 mutation status was assessed from the DNA and RNA of biopsies collected prospectively from 83 patients with lung cancer. Biopsies were obtained either by conventional bronchoscopy or computed tomography-guided percutaneous biopsy. Matched surgical specimens were available for 22 patients. Three assays were used: direct sequencing; a functional assay in yeast; and a newly developed PCR/ligase detection reaction/Universal DNA array assay.

Results: Using the functional assay, p53 mutation was found in 62% of biopsies and 64% of surgical specimens with a concordance of 80%. The sensitivity of the functional assay was determined to be 5%. Direct sequencing confirmed mutations in 92% of surgical specimens but in only 78% of biopsies. The DNA array confirmed 100% of mutations in both biopsies and surgical specimens. Using this newly developed DNA array, we demonstrate the feasibility of directly identifying p53 mutations in clinical samples containing <5% of tumor cells.

Conclusions: The versatility and sensitivity of this new array assay should allow additional development of mutation profiling arrays that could be applied to biological samples with a low tumor cell content such as bronchial aspirates, bronchoalveolar lavage fluid, or serum.

INTRODUCTION

Over the past 20 years, lung cancer has remained the leading cause of cancer-related deaths in the world, and the overall 5-year survival has remained unchanged over this time at an abysmal 15% (1, 2). At present, clinical prognostic indicators such as Tumor-Node-Metastasis staging classification or performance status remain the main parameters used for treatment decisions. A major obstacle to curative treatment of lung cancer is the early onset of extrapulmonary dissemination. Small cell lung cancers are almost never accessible to surgical resection, whereas only 20-30% of non-small cell lung cancer patients presenting with apparently localized disease receive either surgery as sole treatment or multimodality treatment, including chemotherapy and/or radiotherapy with surgery (3).

Lung cancer is the clinical expression of a disease representing the end point of a series of specific somatic genetic and epigenetic changes that precede the invasive tumor by many years (4). These changes include loss of heterozygosity at chromosomes 3p, 9p, 17p, microsatellite instability, p16, and other tumor suppressor gene promoter methylation, K-ras, and/or p53 mutations. The use of these changes as a clonal marker to detect rare tumor cells in body fluids such as sputum, bronchoalveolar lavage, bronchial aspirates, biopsy, and serum would be very promising for the early diagnosis of lung cancers. However to date, the potential prognostic, predictive, and therapeutic value of detecting these alterations has been disappointing, partly due to the lack of power of a single alteration and partly due to heterogeneity between the various assays. Furthermore, many of the studies performed to date have been retrospective, using either frozen tissue or paraffin-embedded samples from surgical specimens. The use of these surgical specimens to screen for new molecular markers in either retrospective or prospective studies may be unintentionally biased because it tends to focus

on only a subset of patients because: (*a*) most lung cancers are unresectable, (*b*) patients with resectable tumors have a better prognosis, and (*c*) patients with resectable cancer generally receive neoadjuvant chemotherapy before surgery.

To meet the challenge of molecular profiling of tumors, there is an urgent need to develop routine molecular diagnostic procedures to manage small or heterogeneous samples such as biopsies, bronchial aspirates, bronchoalveolar lavage, or sputum. It is equally urgent to develop sensitive assays able to overcome the small size and low percentage of tumor cell content of these samples. Biopsies are a suitable material because they are routinely performed in every patient suspected to have a lung tumor.

Among the various potential markers, accurate detection of p53 mutations could be clinically meaningful because this protein plays a key role in drug-induced apoptosis. Consequently, p53 mutational status could influence tumor response to chemotherapy. Furthermore, p53 mutations are frequent and occur early in lung cancer, making them attractive as markers for early detection of tumor cells. The discordance in the literature concerning the clinical relevance of p53 mutational status may be partly caused by different methods of analysis (5). We have recently established that the analysis of the central region of the gene (exons 5–8) misses ~13% of mutations, with half of these mutations corresponding to null mutations (5). The correlation between *p53* gene mutation and p53 protein accumulation in tumor cells is also only 70% based on studies analyzing the entire *p53* gene. This indicates that immunohistochemical analysis is not sufficiently sensitive. Moreover, recent studies have emphasized the concept that p53 mutants may present a heterogeneous behavior. Only a specific subset of p53 mutations could be of clinical value, and this subset could be different depending on the type of cancer or the treatment regimen used (6–11).

We have developed a prospective program to establish routine DNA and RNA extraction of biopsy specimen at the time of diagnosis. In this prospective study, we analyzed the p53 gene status using two sensitive methodologies: the yeast functional assay originally developed by Dr Richard Iggo (12) and the PCR/ligase detection reaction (LDR)/Universal array developed by Dr Francis Barany (13–15). We demonstrate that the yeast assay is more sensitive than direct sequencing for detection of p53 mutations in clinical specimens contaminated by a high proportion of stromal cells and can be used for routine analysis. Use of the PCR/LDR/Universal array also achieves a throughput and sensitivity that cannot be achieved by other currently available technologies.

MATERIALS AND METHODS

Patients. A cohort of 210 consecutive patients was prospectively evaluated for newly suspected lung cancer over a 20-month period (June 2000 to February 2002) in our chest surgery department. Fiber optic bronchoscopy was performed in all patients. Nonsurgical biopsies were used as the diagnostic procedure in 170 patients. Diagnostic material was obtained either by biopsy of an endobronchial lesion visualized during bronchoscopy or by computed tomography (CT)-guided percutaneous biopsy when bronchoscopy was not contributive. During bronchoscopy, four biopsies were taken and fixed in alcohol,

formalin, and acetic acid for diagnosis, and two additional biopsies were taken and snap-frozen in individual cryotubes in liquid nitrogen at the time of endoscopy when the procedure was well tolerated (without respiratory intolerance, excessive cough, or bronchial bleeding). For CT-guided percutaneous biopsy, only one sample was taken and fixed in alcohol, formalin, and acetic acid, and a second biopsy was taken and snap-frozen at the time of CT scan, if well tolerated by the patient. No additional biopsy was performed for the purpose of this study, and all alcohol, formalin, and acetic acid-fixed and snap-frozen-paired biopsies were archived in the Tenon Hospital pathology department. Among the 134 patients from whom snap-frozen biopsies were obtained, the diagnosis of lung cancer could not be performed on alcohol, formalin, and acetic acid-fixed specimens in 28 cases, and the snap-frozen-paired biopsies were used to avoid another diagnostic procedure for the patient. Finally, frozen tissues from 106 patients (86 obtained by bronchoscopy and 26 obtained by CT-guided percutaneous biopsy) were the subject of the present study.

This procedure did not increase the number of biopsies for investigative purposes and only used specimens already acquired for routine diagnosis, as recommended by the French governmental Agence Nationale d'Accréditation et d'Evaluation en Santé in its "Recommendations for tumor cryopreserved cell and tissue libraries for molecular analyses."[5] As recommended, patients were informed that a part of the pathological specimens could be used for molecular analysis provided that a definitive pathological diagnosis was obtained on formalin-fixed samples.

Pathological Procedure. Snap-frozen biopsies, 1–3 μl in diameter and stored at −80°C, were cut in a cryostat chilled to −30°C. To avoid cross-contamination between tissues, the razor was moved 0.5 cm after each section was cut. In this way, a cryostat razor was used to cut 10–12 different specimens. After use, the razor was washed with distilled water, ethanol dried, and exposed for 30 min to a UV bank before starting a new series of sections. A first 5-μm slide was processed with Toluidine blue stain to assess the tumor cell content (Supplementary Figs. 1–7). If the slide contained at least 10% tumor cells, 10–20 adjacent 10-μm frozen sections were cut and immediately placed in a cryotube immersed in liquid nitrogen. Another slide was stained to check that the block still contained tumor cells. If the first frozen section slide did not contain tumor, a second or third section was cut deeper into the tissue block, and frozen slides were only prepared for molecular analysis if this microscopic examination showed the presence of tumor. If three consecutive Toluidine blue stain-stained slides were negative, the sample was not used, and the second frozen sample was accessed for similar processing. Among the 106 biopsies processed, 20 were eliminated because the biopsy was histologically negative for tumor cells, one was eliminated because it corresponded to a lung metastasis from a primary breast cancer, and 2 were eliminated because the tissue was too necrotic. A total of 83 samples was therefore processed for molecular analysis (Table 1). For 22 patients from whom biopsies

---

[5] Internet address: http://www.anaes.fr/ANAES/SiteWeb.nsf?wRubniques!D/APEH-3ZMHJP.

Table 1 Patient characteristics

Characteristics	Total patients (%)	p53 mutation
Age at diagnosis (yrs)		
<60	35 (42 2)	19
>60	48 (57 8)	32
(mean ± SD, range)	(60.8 ± 11 5, 19–82)	
Gender (M/F)	67 (80 7)/16 (19 3)	41/10
Histology		
Non-small cell lung cancer	65 (77)	35
Adenocarcinoma	21" (25)	10"
Squamous cell carcinoma	33 (39)	21
Large cell carcinoma	10 (12)	4
Typical carcinoid	1 (1.2)	0
Small cell lung cancer	19 (23)" (16 9)	17'
Smoking (mean ± SD, range)	(49 6 ± 27; 0–137)	
>30	69 (83 1)	43
<30	10 (12)	6
0	4 (4 9)	2
Disease extent		
Non-small cell lung cancer	65	35
IIIB-IV	32	17
I/II/IIIA	33'	19"
Small cell lung cancer	19	17
Localized	5"	4"
Disseminated	14	13
Total no. of patients	83	51

" IADC + small cell lung cancer (mixed)

were available, surgical specimens were also available leading to a total of 105 samples. The pathologist obtained the samples within 40–60 min after devascularization of the lobectomy or pneumonectomy. Histological control and sectioning were performed as described above. The pathologist (M Antoine) classified these specimens semiquantitatively: + if it contained 0–25% of tumor cells; + + if it contained 25–50%; + + + if it contained 50–75%; and + + + + if it contained 75–100%. The WHO international histological classification was used to assess the final pathological diagnosis. Specimens from 83 subjects were therefore studied in the present article.

Nucleic Acid Extraction and Processing. DNA and RNA extraction was performed simultaneously using the DNA/RNA minikit (Qiagen 14123). Genetic material from surgical specimens was resuspended in either TE [10 mM Tris (pH 8.0), 1 mM EDTA] (DNA) or water (RNA) in a final volume of 20 and 25 μl, respectively. Genetic material from biopsies was resuspended in a final volume of 10 μl. The yield of RNA and DNA allowed multiple independent PCR amplifications for either direct sequencing or functional p53 assay

Reverse Transcriptase-PCR and PCR Analysis. Reverse transcription of RNA was performed using 2 μl of RNA. The RNA was incubated for 5 min at 65°C before adding 18 units of random primers (Invitrogen), 100 units of the Superscript II reverse transcriptase (Invitrogen), 10 mM DTT, 40 units of the RNase inhibitor, RNaseOUT, and 1 25 mM deoxynucleoside triphosphate The reaction was incubated for 1 h at 45°C in a final volume of 20 μl After inactivation at 72°C for 3 min, 2 μl of the cDNA preparation were used for PCR in a final volume of 20 μl [1 25 units of error-free PfU polymerase (Stratagene), 0.5 μM of each primer, 50 μM deoxynucleoside triphosphate, and 10% DMSO] The amplification conditions were as follows 5 min at 94°C, then 30 cycles of 30 s at 94°C,

30 s at 62°C, 2 min at 74°C, followed by 10 min at 74°C (final extension step) Five μl of the product were used for agarose gel analysis. For the yeast assay, the 5'- and 3'-region of p53 cDNA was amplified separately For the 5'-region, we used phosphorothioate-modified primers P3 (ATTTGATGCTGTCCCCG-GACGATATTGAAsC, where s represents a phosphorothioate linkage) and P17 (GCCGCCCATGCAGGAACTGTTACA-CAsT). For the 3'-part, we used P16 (GCGATGGTCTGGC-CCCTCCTCAGCATCTTsA) and P4 (ACCCTTTTTGGACT-TCAGGTGGCTGGAGTsG). The size of these two reverse transcriptase-PCR products was 611 and 569 bp, respectively. For genomic DNA analysis, PCR was performed in a final volume of 25 μl [0.625 units of TaqGold polymerase (Applied BS), 0.2 μM of each primers, 200 μM of each deoxynucleoside triphosphate, 4 mM $MgCl_2$] The amplification conditions were as follows: 10 min at 95°C, then 30 s at 95°C, 30 s at 60°C, 60 s at 72°C (35 cycles), and 10 min at 72°C (final extension step) Primers for amplification of genomic DNA have already been described previously (16) Five μl of the product were used for agarose gel analysis DNAs were sequenced using the Big Dye Read reaction terminator kit (PE Biosystems) and an ABI 3100 genetic analyzer according to the manufacturer's instructions

Yeast Assay. Transcriptional activation is the critical biochemical function of p53, which underlies its tumor suppressor activity Mutant p53 proteins fail to activate transcription. A yeast strain (yIG397), defective for adenine synthesis because of a mutation in its endogenous ADE2 gene but containing a second copy of the ADE2 open reading frame controlled by a p53 response promoter, has been developed. Because ADE2-mutant strains grown on low-adenine plates turn red, yIG397 colonies containing mutant p53 are red, whereas colonies containing wild-type p53 are white For the assay, the yeast strain was cotransformed with reverse transcriptase-PCR-amplified p53 and a linearized expression vector p53 cDNA is therefore cloned in the vector *in vivo* by homologous recombination. To minimize mutations introduced during PCR, we used PfU DNA polymerase (Stratagene), a high-fidelity polymerase In the original assay described by Flaman *et al* (12), only one reverse transcriptase-PCR product was amplified and transformed in the recipient yeast The cutoff for mutation was established as >15% red colonies, indicating the presence of a p53 mutation (12) Although >70% of red colonies are usually obtained for tumors with a high tumor DNA content, ambiguous results may be observed for tumors with a lower tumor cell content or with highly heterogeneous tumor cells We and other authors (17–19) have also observed that the background of red colonies (false positive) can be heterogeneous from one sample to another, leading to difficulties defining a precise cutoff value. This heterogeneity was reproducible from one sample to another, suggesting that each sample of genetic material could have an inherent behavior that could be due either to the quality of the starting material, contaminating compounds affecting the processivity of the enzyme or both. Bearing this problem in mind, Waridel *et al* (20) developed a split functional analysis of separated alleles in yeast (FASAY), where the p53 cDNA is amplified into two overlapping PCR fragments that are independently transformed in the recipient yeast with the appropriate vector The first fragment (P3-P17) corresponds to residues 52–236, whereas the second fragment (P4-P16) corresponds to

residues 195–364. Because there is only one mutation/p53 cDNA, the main advantage of this improvement is that one PCR fragment for each sample will lead to background colonies, whereas the other fragment will lead to red colonies if a mutation is present.

Recovery of p53 Plasmids from Yeast and DNA Sequencing. For each sample yielding >15% of red colonies, the pooled plasmid DNA from 10 red yeast colonies was extracted and sequenced to make a final decision concerning mutations. The plasmid DNA was sequenced using the Big Dye Read reaction terminator kit (PE Biosystems) and an ABI 3100 genetic analyzer according to the manufacturer's instructions. For samples with <15% of red colonies, DNA from 10 red colonies was individually sequenced to distinguish true mutations from the background of PCR errors.

PCR-LDR Assay for p53 Mutations. PCR/LDR/Universal Array assays were generally performed as described in Favis et al.[a] and Gerry et al. (14).

p53 exons 5–8 were simultaneously amplified in single-tube reactions. Primer sequences, in 5′- to 3′-orientation, were as follows: exon 5 forward = CTGTTCACTTGTGCCCT-GACTTTC, exon 5 reverse = CCAGCTGCTCACCATCGCT-ATC; exon 6 forward = CCTCTGATTCCTCACTGATTGCT-CTTA, exon 6 reverse = GGCCACTGACAACCACCCTTAAC; exon 7 forward = GCCTCATCTTGGGCCTGTGTTATC; exon 7 reverse = GTGGATGGGTAGTAGTATGGAAGAAATC; exon 8 forward = GGACAGGTAGGACCTGATTTCCTTAC, and exon 8 reverse = CGCTTCTTGTCCTGCTTGCTTAC. To ensure amplification of all exons, PCR was performed by using primers containing a universal primer sequence at the 5′-ends. The initial PCR reaction was performed as previously described (13, 15) with the following modifications. The 25-μl PCR reaction mixture contained 3–5 μl of primary tumor DNA, all four deoxynucleoside triphosphates (400 μM of each one), 10 mм Tris-HCl (pH 8.3), 50 mм KCl, 4 mм MgCl$_2$, 0.625 units of AmpliTaq Gold (PE Applied Biosystems, Inc., Norwalk, CT), 2 pmol of gene-specific primers containing a 5′-universal sequence for exons 5, 6, and 8, and 4 pmol of a similar primer for exon 7. The reaction was preincubated for 10 min at 95°C. Amplification was performed for 15 cycles as follows: 94°C for 15 s and 65°C for 1 min. A second 25-μl aliquot of the reaction mixture, containing 25 pmol of universal primer, was then added. PCR was repeated for 25 cycles at an annealing temperature of 55°C for 1 min. Amplification was verified by examining the products on 3% agarose gel. Taq polymerase was inactivated by 3 cycles of freezing in dry ice.

After a multiplex PCR amplification of the regions of interest, each mutation was simultaneously detected using a thermostable ligase that joins pairs of adjacent oligonucleotides complementary to the sequences of interest. Ligation occurs only when there is perfect complementarity at the junction between the 5′-fluorescent-labeled upstream oligonucleotide, containing the discriminating base for the mutation on the

3′-end, and the adjacent downstream oligonucleotide, containing a complementary zip code sequence on the 3′-end. The complete set of LDR primers is described in Favis et al.[a] Ligation products are distinguished on the basis of differential labeling and capture of the zip code complement on its cognate zip code address on an universal array.

LDR reactions were carried out in a 20-μl mixture containing 20 mм Tris-HCl (pH 7.6), 10 mм MgCl$_2$, 100 mм KCl, 10 mм DTT, 1 mм NAD$^+$, 25 nм (500 fmol) of the detecting primers, 2 μl of PCR product, and 25 fmol of Tth DNA ligase. Ligases were overproduced and purified as described previously (21, 22). LDR reactions were incubated for 5 min at 95°C and were then thermally cycled for 20 cycles of 30 s at 95°C and 4 min at 64°C. Quality control for LDR was performed using a synthetic template for each mutation to test the ability of the full mix of upper or lower ligation primers to produce the expected specific signal on the DNA microarray.

Preparation and hybridization were performed as previously described (13, 14), except that hybridization was carried out in the presence of 100 μg/ml sheared calf thymus DNA. Briefly, 20 μl of the LDR reaction were diluted with 20 μl of 2.0× hybridization buffer to produce a final buffer concentration of 300 mм 4-morpholineethanesulfonic acid (pH 6.0), 10 mм MgCl$_2$, and 0.1% SDS that was incubated for 5 min at 94°C before loading in the chips. The arrays were placed in humidified culture tubes and incubated for 1 h at 65°C and 20 rpm in a rotating hybridization oven. After hybridization, the arrays were washed in 300 mм bicine (pH 8.0), 10 mм MgCl$_2$, and 0.1% SDS for 10 min at 65°C. Arrays were reused three times and were stripped between uses by submerging for 1 min in a solution of boiling 100 mм bicine/0.1% SDS. Stripped arrays were rinsed in nanopure water, excess water was removed using forced air, and the arrays were stored in slide boxes at room temperature.

RESULTS

The clinical and histological characteristics of 83 patients with lung cancer are shown in Table 1. The distribution of the various histological types is in agreement with recent data concerning the distribution of lung cancer in France, indicating that no recruitment bias occurred during this prospective study (23).

Using total RNA extracted from either the biopsy or the tumor sample, reverse transcriptase-PCR amplification and FA-SAY analysis of all 105 samples (100%) were successful (Supplementary Fig. 1). FASAY analysis for the detection of p53 mutations has been extensively described, but most of these studies used a first generation assay with only one PCR product corresponding to residues 52–364. The cutoff value of red colonies for a positive result is usually arbitrarily defined between 10 and 20% (24–26). In the present study, we first used a 15% cutoff value, leading to the detection of p53 mutations in 44 of 83 biopsies and 14 of 22 tumors. Direct sequencing of pooled rescued plasmid DNA from yeast led to the identification of the p53 mutation in 100% of cases (Supplementary Figs. 1–7).

In the split methodology, the *p53* gene is cloned into two fragments. The basic idea is that the number of red colonies arising in the second fragment not containing the *p53* mutation will always correspond to background mutations. Two p53

[a] R. Favis, J. Huang, N. P. Gerry, A. Culliford, P. Paty, T. Soussi, and F. Barany. Harmonized microarray mutation scanning analysis of p53 mutation in undissected colorectal tumors, in press, Human Mutation, June 2004.

mutations are very rarely found in the same allele of the gene. We calculated the mean percentage of red colonies generated by the negative fragment of each tumor bearing a *p53* mutation. Samples with mutations in the overlapping segment of the two PCR products were carefully removed. Only samples with >15% of red colonies were taken into account in this analysis. This statistical analysis of the cutoff values was based on 39 samples of P3-P17 and 32 samples of P4-P16 fragments. The mean percentage of red colonies was 3.4 ± 2.6% for P3-P17 and 4.0 ± 2.4% for P4-P16. Similar mean values were obtained when the same analysis was performed on tumors negative for p53 mutations. Using cutoff values of 8.6 and 8.8% (mean ± 2 SDs), 7 biopsy specimens gave percentages of red colonies ranging between these cutoff values and our previous limit of 15% (Table 2). No new p53 mutations were detected among the surgical specimens. For these 7 specimens, sequencing of 10 individual red colonies led to the detection of p53 mutations (see "Materials and Methods"). The case of B32 is also noteworthy. Petri dishes transformed with the P3-P17 PCR product led to 7.2% of red colonies and 5.1% of pink colonies. These pink colonies have been shown to originate from leaky *p53* mutations that do not completely inactivate *p53* function (25, 26). Sequencing of 10 individual clones from pink colonies detected a single substitution at codon 180 of the p53 gene in a region known to lead to mutant p53 with a mild phenotype, whereas sequencing of individual clones from red colonies led to the identification of multiple mutations arising from PCR amplification. This particular example clearly shows that the split FASAY is a very sensitive method to detect mutant p53 in a highly heterogeneous tumor sample.

Therefore, using the new cutoff value defined above, 52 of 84 biopsies (62%) and 14 of 22 tumors (64%) were positive on the FASAY (Tables 2, 3, and 4).

The spectrum of missense mutations was as follows: 11 (G:C → A·T) transitions, 6 of which occurred at a CpG dinucleotide, 19 (G·C → T·A), 6 (T:A → C·G), 3 (A·T → T·A), and 5 (G·C → C·G) transversions. Nine frameshift mutations and 1 splice mutation were also revealed (Table 2). The high frequency of GC → TA transversions, which are usually only found in lung cancer patients, is associated with tobacco smoking (27). Five mutations were found in the 157–159 region, a hot spot region that has been shown to be the specific target of the tobacco carcinogen benzo(*a*)pyrene (28). The concordance between the pattern of p53 mutations described in this article and published literature based on more conventional procedures indicates that the functional assay used in the present study did not induce any specific selection bias for p53 mutations. This pattern of mutational events is not unexpected because the majority of patients in the present series were smokers (Tables 2 and 3).

In the series of 22 matched samples of biopsies with surgical specimens, 7 samples were wild-type in both samples, 11 had the same mutations, and 4 were discordant (Table 3).

To validate this FASAY analysis, direct sequencing was performed using either DNA or cDNA as starting material. The identity of the p53 mutation was confirmed in 28 of the 39 biopsies (71%) and 12 of the 13 (92%) surgical samples, whereas no mutation was detected in the remaining samples (Tables 2, 3, and 4). It is noteworthy that cDNA sequencing was more sensitive on 3 samples, confirming previous observations

that mutant p53 RNA may be more stable or may be expressed at a higher level in tumor cells (29). Failure of sequencing is certainly caused by the low tumor cell content in the sample and the lack of sensitivity of automatic sequencing.

We have recently developed a microarray-based assay to detect p53 mutations that uses a thermostable ligase enzyme to discriminate between wild-type and mutant templates, resulting in separation of mutation detection and array hybridization (13–15).[6] This assay was used to efficiently detect *p53* mutations in surgical specimens from patients with colorectal cancer, but its sensitivity in nonsurgical samples such as biopsies has not been previously tested. Nine surgical specimens and 27 biopsies with p53 mutations detected by the FASAY were available for analysis by the array (Table 5 and Fig. 1F). The array confirmed mutations in all of the 27 biopsies (100%, Table 5), 7 (27%) of which were not confirmed by direct genomic DNA sequencing (Tables 2 and 3). Two mutations not detected by direct sequencing were also detected by the array. All *p53* mutations were detected by the array for the 8 surgical samples. For patient C6 in whom biopsy and surgical specimens were both available, histological examination of the specimen and FASAY analysis indicated a higher tumor cell content for the surgical specimen (70 versus 30%). Although FASAY easily detected a mutation at codon 249 in both samples, direct sequencing of the biopsy failed to detect the mutation, whereas the DNA chips clearly identified this event (Fig. 2, *A–F*). This feature can be applied to the majority of the samples analyzed in this study and emphasizes the high sensitivity of this array technology for biopsy specimens.

DISCUSSION

Lung carcinomas are typically late-stage and biologically aggressive, which accounts for their poor prognosis (4). The potential of new imaging and molecular techniques to significantly improve the detection of localized lung cancer provides an unprecedented opportunity to understand the biology, improve diagnosis, enhance treatment, and reduce mortality (30). Furthermore, recently developed proteomic and expression array technologies have intensified the search for new biomarkers that could be helpful in defining response to therapy or prognosis.

Only 30% of patients with non-small cell lung cancer and <5% of patients with small cell lung cancer are treated surgically, implying that the biological sample most frequently available for routine management at the time of diagnosis is biopsy. The size and heterogeneity of biopsies raise problems for current molecular diagnosis techniques. There is therefore an urgent need to develop sensitive assays for the detection of lung tumor-specific molecular alterations in routinely available specimens such as biopsies, bronchoalveolar lavage, or sputum. In the present prospective study, we demonstrate the feasibility of routine management and analysis of lung biopsy specimens for p53 mutation. This includes biopsies obtained using conventional bronchoscopy as well as CT-guided percutaneous biopsy. To our knowledge, this is the first time that material obtained by CT-guided percutaneous biopsy has been processed for molecular analysis despite the smaller sample size compared with biopsies obtained by conventional procedures. This is important in view of the increasing worldwide rate of adenocarcinoma in

Table 2 Analysis of lung biopsies for p53 mutations by functional analysis of separated alleles in yeast, direct sequencing, and DNA chips

Sample	Histology	% tumor cells[a]	Functional analysis of separated alleles in yeast[b]		Mutation (FASAY analysis)		DNA analysis[c]	RNA analysis[d]	Chips[e]
			% red clones 5'	% red clones 3'	Codon	Mutational event			
B1[f]	SCLC	100/75	10 8	87 8	249	AGG→AGT	(+)	ND	(+)
B2	SCC	50/75	97 1/96 9	68/71 7	305	TAT→TGT	(+)	ND	(+)
B3	LCC	75/75	64 3	2	132	AAG→AAC	(+)	ND	NA
B4	SCLC	75/75	65.7	3 1	192	CAG→TAG	(+)	ND	NA
B5	SCLC	10/10	86 7	2 4	157	GTC→TTC	(+)	ND	NA
B6	SCC	75/75	21 8	5 7	144	CAG→TAG	(+)	ND	(-)
B7	SCLC	75/100	0/4 4	44.2/18 4	del	Del part exon 8 and intron 8	(+)	ND	(+)
B8	SCLC	25/15	12 1	38 5	220	TAT→TGT	(+)[g]	(+)[g]	NA
B9	SCC	50/50	1 3	38 5	220	TAT→TGT	(+)	ND	(-)
CTB10	SCC	15/0	2 1	60 2	239	AAC→GAC	ND	(+)	NA
B11	NSCLC	10/0	22 5	91 8	267	COG→CCG	(+)	ND	NA
B12	SCLC	10/50	13 9	2 2	110	CGT→CTT	(-)	(-)	NA
B13	ADC	75/50	81 9	63	249	AGG→TGG	(+)	ND	(-)
B14	SCC	25/E	83	33 6	195	ATC→AAC	(+)	ND	NA
B15	SCLC	100/75	53 9	2 5	175	CGC→CAC	(+)	(+)	(-)
B16	ADC	0/5	76 6	5 4	100	CAG→TAG	(+)	ND	NA
B17	SCLC	75/100	58 6	5 3	157	GTC→TTC	(-)	Weak	(+)[h]
B18	LCC	75/75	1 8	82 1	220	TAT→TGT	(+)	ND	(+)
B19	SCLC	75/75	1 5	97 1	300	DEL C1	(+)	ND	NA
B20	ADC	10/10	2 5/2 3	64	273	CGT→CTT	(+)	ND	(-)
B21	SCC	100/100	1 5	9 3/9 4	306	CGA→TGA	(-)	ND	(-)
B22	SCC	100/75	3 8/2 2/7 2	94 9	273	CGT→CTT	(+)	ND	(-)
B23	ADC	75/75	82 6	8 2. 15 6/10 1	278-79	T insertion	ND	ND	NA
B24	SCC	75/75	69 2/64 8	1 8	179	CAT→CGT	(+)	ND	(-)
B25	SCC	50/75	1 9	2 5 16	110	CGT→CTT	(+)	ND	NA
B26	SCC	50/0	1 1	84 2	245	GGC→TGC	(+)	ND	(-)
B27	SCLC	75/50	2 1	58 2	278	CCT→TCT	(-)	ND	(-)
B28	SCLC	75/100	10 9 16 1	64.6	286	DEL G1 (GGA)	ND	ND	NA
B29	ADC	10/E	2 8	4.1 80 8	237	ATG→AAG	ND	ND	NA
B30	ADC	75/100	5 2	31 3	373	CGT→CAT	(+)	ND	(-)
B31	ADC	75/75	10	41 7	298 99	17 bp insertion	ND	ND	NA
B32	SCLC	10/10	7 2 5 1	5 9	134 35	TTTGC→TTCACC	ND	ND	NA
B33	ADC	25/50	11 6	3 4	180	GAG→GAT	ND	ND	NA
B34	SCLC	75/75	9 5 7 2	1 4	157	DEL C3 (GTC)	ND	ND	NA
B35	SCLC	75-75	65 6	49 6/49 5	307	DEL G1 (GCA)	ND	ND	NA
B36	ADC	50/15	13 4/15 4	4 7	179	CAT→AAT	ND	ND	NA
B37	ADC	10/15	6 1	6.8/8 7	175	CGC→CAC	(-)	ND	(-)
B38	SCLC	100/100	15 6	30 5	237	ATG→ATT	(-)	ND	(-)
B39	SCLC	100/100	47 3	8	183	TCA→TGA	ND	ND	NA
B40	SCC	50/75	0 7/3 3	0	192	CAG→TAG	ND	ND	NA
B41	ADC	10/10	8 1/6 7	2 8/2 5	WT				
B42	SCC	0/10	0 9/8 5	4/5 2	WT				
B43	SCC	15/15	3 8/7 9	4 6/5 4	WT				
B44	ADC	15/25	0 9/3 3	3 2/6 1	WT				
CTB45	ADC	75/50	2 6/6 4	1 9/2 5	WT				
B46	SCC	0/0	4 6/1 3	6 8/8 4	WT				
B48	NSCLC	5/5	3 6/3	5 8/4 2	WT				
B49	NSCLC	10/0	3 6/4 3	6 1/7 3	WT				
CTB50	ADC	50/25	2 9/2 5	6/7.9	WT				
B51	NSCLC	15/10	6 7	3 1/2 4	WT				
B52	SCLC	0/0	3	2 1	WT				
B53	SCC	50/75	3 7/4 8	3 1	WT				
B54	SCLC	100/100	3 8/4 1	6 4/5 3	WT				
B55	ADC	10/15	1 4/2 4	3 8/3 5	WT				
B56	NSCLC	25/25	1 5/4 4	2 7/3 9	WT				
CTB57	NSCLC	10/0	4 7/1	4 6/5 4	WT				
CTB58	NSCLC	10/E	4.5/4 5	2 3/5 2	WT				
CTB59	SCC	10/E	0 8/8 1	3 4/6 7	WT				
B60	ADC	10/10	7 6/6 8	7 3/2 9	WT				
B61	SCC	10/10	2 3/0 7	3/1 5	WT				
B62	ADC	50.50	5/2 6	2 9/0 8	WT				

[a] The two values correspond to the top and bottom slides, respectively

[b] Frequency of red clones is given for the 5'-part (P3-P17) and 3'-part of p53 (P4-P16) More than 1 assay was performed in several experiments, and all results are shown

[c] Detection of p53 mutation by direct DNA sequencing of genomic DNA +, the same mutation was detected in DNA, −, no mutation detected

[d] Detection of p53 mutation by direct DNA sequencing of cDNA. +, the same mutation was detected in cDNA, −, no mutation detected

[e] Chip analysis was always performed with genomic DNA, except for a few cases in which it was performed with cDNA

[f] B, biopsy obtained by conventional bronchoscopy. SCLC small cell lung cancer, ND, not done. SCC, squamous cell carcinoma, LCC, large cell carcinoma, NA, the mutation is not available on the chip. CTB, computed tomography-guided percutaneous biopsy, ADC, adenocarcinoma, E, block exhausted. WT, wild type, NSCLC, non small cell lung cancer, CT, carcinoid tumors.

[g] Mutation described previously (30)

[h] Signal obtained with cDNA amplified from the tumor. No signal was obtained with genomic DNA

[i] Leaky mutations leading to both red clones (first number) and pink clones (second number)

[j] Nonsmoking patient

Table 3   Analysis of biopsies and surgical specimens from matched patients for *p53* mutations by functional analysis of separated alleles in yeast, direct sequencing, and DNA chips

Sample[a]	Histology	% tumor cells[b]	Functional analysis of separated alleles in yeast[c]		Mutation (FASAY analysis)		DNA analysis[d]	RNA analysis[e]	Chips[f]
			% red clones 5'	% red clones 3'	Codon	Mutational event			
C1B	SCC[g]	10/20	6.7	15.8	275	TGT→TTT	(−)	ND	(+)
C1T	SCC	100/100	0	74.4	275	TGT→TTT	(−)	(+)	(+)
C2B[h]	SCC	0/0	6.6/7.2	18.6/10.9	WT		ND	ND	(−)
C2T	SCC + ADC	75/75	1.7/1.3	48.4/40.6	273	CGT→CTT	(−)	ND	(+)
C3B	SCC	100/100	12.8/11.2	4.3/1.4	71	DEL C1 (CCC)	(+)	(+)	NA[i]
C3T	SCC	100/100	19.3/26.9	1.7/2.1	71	DEL C1 (CCC)	(+)	ND	NA
C4B	SCC	75/100	2.3	92.1	242	TGC→TTC	(+)	ND	(+)
C4T	SCC	100/100	3.3	86.9	242	TGC→TTC	(+)	(+)	(+)
C5CTB	ADC	15/20	5.1/5.8	10.6/11.9	224	GAG→GTCTG	(+)	ND	NA
C5T	ADC	75/75	2.8/5	15.8/12.7	224	GAG→GTCTG	(+)	ND	NA
C6CTB	NSCLC	5/N	0.7	9.9	249	AGG→ATG	(−)	(−)	(+)
C6T	ADC	75/75	4.5	26.6	249	AGG→ATG	(−)	(+)	(+)
C7B	SCC	50/10	1.3	30.7	273	CGT→CAT	(−)	ND	(+)
C7T[i]	SCC	75/75	0.7	66.9	273	CGT→CAT	(+)	ND	(+)
C8B	SCC	100/100	1.7	78.9	273	CGT→GGT	(+)	ND	NA
C8T	SCC	100/100	1.4	82.9	273	CGT→GGT	(+)	ND	NA
C9CTB	NSCLC	75/75	52.2	2.3	159	GCC→CCC	(+)	(+)	(+)
C9T	ADC	100/100	51	2.5	159	GCC→CCC	(+)	(+)	(+)
C10B	SCC	5/0	18.3/43.2	7/10.3	175	CGC→CAC	(−)	ND	(+)[i]
C10T[i]	SCC	50/0	52.7/44.2	2.7/0	105	GGC→TGC	(−)	ND	NA
C11B	ADC + SCLC	100/100	3.5	60.3	248	CGG→CTG	(+)	ND	(+)
C11T[i,k]	SCLC	25/25	4.1	6.1	WT		ND	ND	(−)
C12B	SCC	75/50	44.5	98.8	218–221	DEL 9 PB	ND	ND	NA
C12T	SCC	75/75	30.5	67.1	218–221	DEL 9 PB	(+)	ND	NA
C13CTB	ADC	20/10	21.1	4.6	193	CAT→CGT	ND	ND	(+)
C13T	ADC	75/75	44.3	3	193	CAT→CGT	ND	ND	(+)
C14B[k]	SCC	0/5	6.2/6.5/9.1	8.5/9.7/4.9	WT		(+)	ND	(+)
C14T[i]	SCC	75/100	6.4	71.4	WT		ND	ND	(−)
C15CTB	ADC	50/25	68.7	3.8	248	CGG→CTG	(+)	ND	(+)
C15T[i]	ADC	100/100	77.4	1.5	158	CGC→CTC	ND	ND	(+)
C16B	Ca	75/75	0.2	2.7	158	CGC→CTC	(+)	ND	(+)
C16T	Ca	100/100	1.3	2.5	WT				
C17CTB	SCC	15/25	1.5	2.7	WT				
C17T	SCC	75/75	0.9	1.7	WT				
C18CTB	SCC	0/5	6.2	0.8	WT				
C18T	SCC	100/100	3.4	8	WT				
C19B	SCC	25/25	1.6	4.8	WT				
C19T	SCC	75/75	1.4	2.4	WT				
C20CTB	SCC	75/75	1.1	3.2	WT				
C20T	SCC	100/75	1.8	2.9	WT				
C21CTB	ADC	75/25	5.6	8.3	WT				
C21T	ADC	50/ND	8.6	3.7	WT				
C22CTB	ADC	15/5	1.6	2.9	WT				
C22T	ADC	50/75	0.3	1.3	WT				

[a] Matched biopsies (top lane, suffix B or CTB as defined in Table 2) and surgical specimens (bottom lane, suffix T). All patients were smokers.
[b] The two values correspond to the top and bottom slides, respectively.
[c] Frequency of red clones is given for the 5'-part (P3-P17) and 3'-part of p53 (P4-P16). More than 1 assay was performed in several experiments and all results are shown.
[d] Detection of p53 mutation by direct DNA sequencing of genomic DNA. +, the same mutation was detected in DNA; −, no mutation detected.
[e] Detection of p53 mutation by direct DNA sequencing of cDNA. +, the same mutation was detected in cDNA; −, no mutation detected.
[f] Chip analysis was always performed with genomic DNA, except for a few cases in which it was performed with cDNA.
[g] SCC, squamous cell carcinoma; ND, not done; WT, wild type; ADC, adenocarcinoma; NA, the mutation is not available on the chip; NSCLC, non-small cell lung cancer; N, necrosis; SCLC, small cell lung cancer; Ca, carcinoid tumors; T, surgical specimen.
[h] The discrepancy between the surgical specimen and the biopsy could be due to the very low tumor cell content of the biopsy.
[i] Signal obtained with cDNA amplified from the tumor. No signal was obtained with genomic DNA.
[j] These patients received neoadjuvant chemotherapy.
[k] Histological examination of the biopsy detected a composite tumor consisting of SCLC and adenocarcinoma tissue. Histological examination of the surgical specimen after treatment showed only the SCLC component with WT p53, suggesting that the p53 mutation observed in the biopsy could arise from the adenocarcinoma component.

Table 4 Summary of p53 mutation analysis

| | Single biopsies | Matched biopsies/tumors | |
		Biopsies	Tumors
FASAY[a]	39/62 (63%)	13/22 (60%)[b]	14/22 (63%)[b]
Sequencing[c]	22/28 (78%)	7/11 (63%)	12/13 (92%)
Chips[d]	18/18 (100%)	9/9 (100%)	9/9 (100%)

[a] Functional analysis of separated alleles in yeast (FASAY) represents the true frequency of p53 mutations in this series because no patient selection was performed for the analysis.

[b] Two patients were negative for the biopsies but positive for the tumor: 1 patient has a different mutation in the tumor and in the biopsy, and 1 patient with a mixed tumors (small cell lung cancer + non-small cell lung cancer) had a positive biopsy and a negative tumor (see text for more details)

[c] Only patients with positive FASAY are indicated. No p53 mutation was found in negative patient (see text for detail)

[d] Only patients with a p53 mutation and for whom the chips assay was available were tested

Table 5 DNA chip analysis[a]

	SEQ +/ ARRAY +	SEQ −/ ARRAY +	SEQ ND/ ARRAY +	SEQ +/ ARRAY −
Tumors	9	0	0	0
Biopsies	17	8	2	0

[a] All samples analyzed by the array were shown to contain a p53 mutation after by functional analysis of separated alleles in yeast analysis. SEQ, detection of p53 mutation by direct sequencing; ARRAY, detection of p53 mutation by PCR/LDR array

which CT-guided percutaneous biopsy is the method of choice for these peripheral tumors.

Although p53 mutations are common in lung cancer, the importance of these mutations for the patient's clinical outcome is still controversial (5), mainly because of the heterogeneous strategies used to assess p53 mutational status. Immunostaining lacks sensitivity because of false negatives from nonsense mutations, splicing mutations, and deletions that do not lead to p53 accumulation. In the present study, 10 mutations could not have been detected by immunostaining, and the splice mutation could not be detected by DNA sequencing (31). The majority of molecular analyses have also focused on the study of p53 exons 5 through 8. In a recent analysis of the p53 mutation database, we showed that this bias results in nondetection of ~13% of p53 mutations, and these false negatives may bias interpretation of the results during statistical analysis.

In the present study, we compared assays based on either DNA, direct sequencing or arrays, or RNA, the functional assay in yeast. Initially developed for the detection of germ-line mutations, the yeast assay has been widely used for the detection of somatic mutations in various types of tumors, including a few studies in lung cancer (32, 33). The yeast assay can be used to screen p53 from exons 4 to 10, which accounts for >95% of p53 mutations. In the present study, using the new split assay developed by Waridel et al (20) and an experimentally defined cutoff value, we show that this assay may be sufficiently sensitive to detect p53 mutations in samples containing only 5% of tumor cells. Sequencing of rescued plasmids from red colonies allowed unambiguous identification of p53 mutations in all cases, but direct sequencing of genomic DNA was only able to detect 72% of mutations in biopsy specimens. Until a more sensitive and specific methodology has been developed, we believe that the yeast assay should be considered as a reference method for the evaluation of p53 mutations in clinical specimens, especially specimens with a low tumor cell content. In addition to the advantages described above, the FASAY can easily distinguish true inactivating mutations from neutral mutations. Furthermore, the use of a short amplicon in reverse transcriptase-PCR also allows this assay to be performed on biological samples that could lead to extraction of partially degraded RNA (19).

Although sensitive, this assay has two major drawbacks: it has a low throughput and it does not provide any information about the precise p53 mutation, therefore, requiring sequencing of rescued plasmids. Although the first limitation could be circumvented by automation, the second limitation could be particularly inconvenient in view of the markedly heterogeneous behavior of various p53 mutants, leading to different clinical phenotypes. Several studies in breast cancer suggest that only specific p53 mutations are associated with de novo resistance to doxorubicin (9).

The PCR/LDR/Universal array assay provides both high throughput and allows direct identification of the mutational event, a feature that considerably reduces the cost of this assay. Furthermore, as demonstrated in the present work, it has a higher sensitivity than direct sequencing. One of the most useful aspects of the PCR/LDR/Universal array is its versatility because the same array can be used for the detection of mutations in multiple genes such as p53, APC, K-ras, or BRCA1 (13, 14). Our laboratories are also developing the PCR/LDR/Universal array to monitor gene promoter hypermethylation,[7] which is a frequent event in various types of cancer, including lung cancer (34, 35). Belinsky et al (36) measured hypermethylation of the CpG islands in the sputum of lung cancer patients and demonstrated a high correlation with early stages of non-small cell lung cancer, which indicated that p16 CpG hypermethylation could be useful in predicting future lung cancer.

We envision the practical development of very sensitive PCR/LDR/Universal array assays, specifically programmed to a given type of cancer such as lung or colon cancer. By querying specific genes for each type of cancer (e.g., gene mutations or hypermethylation), it would be possible to achieve a specificity of 90–95% for identification of tumor cells. Such universal array assays will be very useful to assess the tumor content of clinical specimens such as stool, serum, bronchoalveolar lavage fluid, and sputum— samples that are known to have a low tumor cell content. Using a new standardized extraction and conservation protocol, we have been able to extract RNA and DNA from bronchial secretions aspirated during fiber-optic bronchoscopy (bronchial aspirates) that are considered to contain tumor cells. FASAY and chips analysis were successfully performed with this material, indicating the feasibility of this type of analysis on heterogeneous specimens.[8]

Although the specificity of each gene queried is not high (current chips are programmed to detect only 50% of p53

[7] Y-W Cheng and F Barany, unpublished observations
[8] C. Fouquet, M Antoine, N Rabbe, J Cadranel, G Zalcman, and T Soussi, unpublished results.
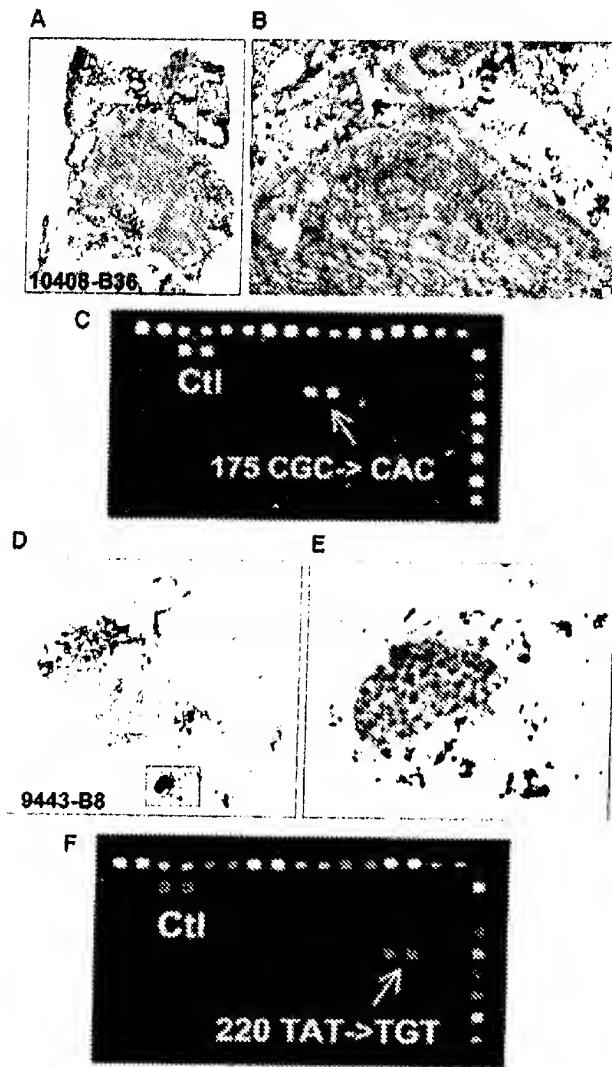
Fig. 1 Histology and array analysis of two bronchial biopsies *A*, *B*, *D*, and *E*, Toluidine blue staining of an adenocarcinoma (*A* and *B*) and a small cell lung cancer (*D* and *E*) *A* and *D*, ×25, *B* and *D*, ×100 *C* and *F*, results of PCR/ligase detection reaction/Universal DNA microarray analysis of DNA Addresses are double spotted onto a three-dimensional surface comprised of a loosely cross-linked polymer of acrylamide and acrylic acid. The three-dimensional surface combined with the zip code system allows hybridized arrays to be stripped of target and reused Fiducials labeled with Cy3 Bodipy, and Alexa are spotted along the *top* and the *right side* of the array to provide orientation. Amplicon controls (Ctl) are seen in the next row, the Cy3 signal indicates that samples 10408 and 9443 present 175 G→A and 220 A→G mutations, respectively

mutations in lung tumors), the probability of finding an index marker among the multiple genes queried is very high. The use of multiple fluorochromes could also improve the throughput of the assay.[9]

Many small and early lesions are now being detected in high-risk individuals by either low-dose CT scan screening programs or endoscopic fluorescence devices, but their true clinical significance remains uncertain. It is not possible to predict which of these lesions will really progress toward either overt cancer for dysplastic bronchial epithelial lesions or metastatic disease for early-stage cancers. It may be appropriate to target these premalignant changes or small stage I tumors for
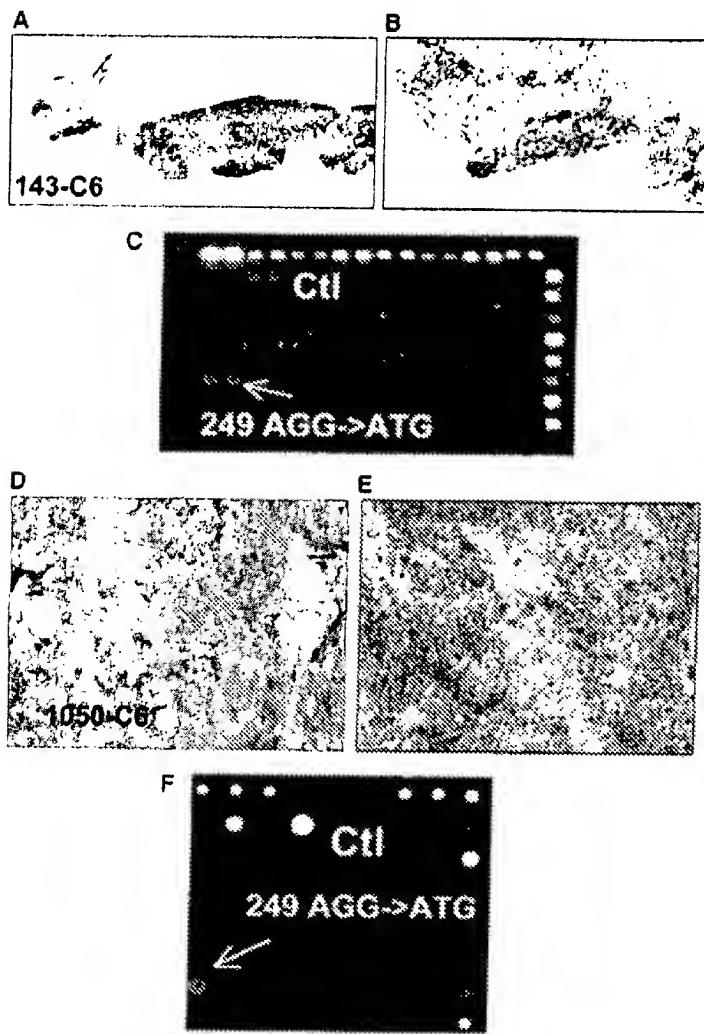
Fig. 2 Histology and array analysis of a matched biopsy and surgical specimen from the same patient. Toluidine blue staining of the biopsy (*A* and *B*) and surgical specimen (*D* and *E*) at two magnifications. *A* and *D* (×25), *B* and *D* (×100). *C* and *F*, results of PCR/ligase detection reaction/Universal DNA microarray analysis of DNA. Amplicon controls (Ctl) are seen in the *top row*; both samples display the same G→T mutation at codon 249. The arrangement of capture oligonucleotides in the array displayed in *F* is different because of a new spotting procedure.

early detection and intervention by fully profiling their molecular characteristics, including evaluation of response to specifically targeted intervention. High-throughput technologies such as genomics and proteomics are becoming widely available, and it will be crucial to apply these technologies to the detection of early lung carcinogenesis and outcome assessment. However, all of these technologies, including sample management and extraction of nucleic acids, must also be feasible as routine procedures in major clinical departments. The data presented here suggest that the PCR/LDR/Universal array assay, applied to samples containing a minority of tumor cells or DNA, recruited prospectively, meets these requirements.

ACKNOWLEDGMENTS

REFERENCES

1. Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics, 1999. CA - Cancer J Clin 1999;49:8–31.

2. Tyczynski JE, Bray F, Parkin DM. Lung cancer in Europe in 2000: epidemiology, prevention, and early detection. Lancet Oncol 2003;4:45–55.

3. Mountain CF. Revisions in the international system for staging lung cancer. Chest 1997;111:1710–7.

4. Hirsch FR, Franklin WA, Gazdar AF, Bunn PA Jr. Early detection of lung cancer: clinical perspectives of recent advances in biology and radiology. Clin Cancer Res 2001;7:5–22.

5. Soussi T, Béroud C. Assessing TP53 status in human tumours to evaluate clinical outcome. Nat Rev Cancer 2001;1:233–40.

6. Huang C, Taki T, Adachi M, Konishi T, Higashiyama M, Miyake M. Mutations in exon 7 and 8 of p53 as poor prognostic factors in patients with non-small cell lung cancer. Oncogene 1998,16,2469–77.

7. Skaug V, Ryberg D, Kure EH, et al. p53 mutations in defined structural and functional domains are related to poor clinical outcome in non-small cell lung cancer patients. Clin Cancer Res 2000,6:1031–7.

8. Tomizawa Y, Kohno T, Fujita T, et al. Correlation between the status of the p53 gene and survival in patients with stage I non-small cell lung carcinoma. Oncogene 1999,18 1007–14.

9. Aas T, Borresen AL, Geisler S, et al. Specific p53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. Nat Med 1996;2:811–4.

10. Borresen AL, Andersen TI, Eyfjord JE, et al. TP53 mutations and breast cancer prognosis. Particularly poor survival rates for cases with mutations in the zinc-binding domains. Gene Chromosome Cancer 1995,14:71–5.

11. Erber R, Conradt C, Homann N, et al. TP53 DNA contact mutations are selectively associated with allelic loss and have a strong clinical impact in head and neck cancer. Oncogene 1998,16:1671–9.

12. Flaman JM, Frebourg T, Moreau V, et al. A simple p53 functional assay for screening cell lines, blood, and tumors. Proc Natl Acad Sci USA 1995;92:3963–7.

13. Favis R, Day JP, Gerry NP, Phelan C, Narod S, Barany F. Universal DNA array detection of small insertions and deletions in BRCA1 and BRCA2. Nat Biotechnol 2000;18:561–4.

14. Gerry NP, Witowski NE, Day J, Hammer RP, Barany G, Barany F. Universal DNA microarray method for multiplex detection of low abundance point mutations. J Mol Biol 1999,292 251–62.

15. Dong SM, Traverso G, Johnson C, et al. Detecting colorectal cancer in stool with the use of multiple genetic targets. J Natl Cancer Inst (Bethesda) 2001,93 858–65.

16. Lidereau R, Soussi T. Absence of p53 germ-line mutations in bilateral breast cancer patients. Hum Genet 1992;89.250–2.

17. Chappuis PO, Estreicher A, Dieterich B, et al. Prognostic significance of p53 mutation in breast cancer: frequent detection of non-missense mutations by yeast functional assay. Int J Cancer 1999;84:587–93.

18. Bonnefoi H, Ducraux A, Movarekhi S. et al. p53 as a potential predictive factor of response to chemotherapy. feasibility of p53 assessment using a functional test in yeast from trucut biopsies in breast cancer patients. Br J Cancer 2002,86 750–5.

19. Tisserand P, Fouquet C, Marck V, Mallard C, Fabre M, Vielh P, Soussi T. ThinPrep-processed fine-needle samples of breast are an effective material for RNA- and DNA-based molecular diagnosis application to p53 mutation analysis. Cancer 2003,99 223–32.

20. Waridel F, Estreicher A, Bron L, et al. Field cancerisation and polyclonal p53 mutation in the upper aerodigestive tract. Oncogene 1997;14:163–9.

21. Barany F, Gelfand DH. Cloning, overexpression and nucleotide sequence of a thermostable DNA ligase-encoding gene. Gene (Amst) 1991,109 1–11.

22. Luo J, Bergstrom DE, Barany F. Improving the fidelity of Thermus thermophilus DNA ligase. Nucleic Acids Res 1996,24 3071–8.

23. Blanchon F, Grivaux M, Collon T, et al. Epidemiologic of primary bronchial carcinoma management in the general French hospital centers [in French]. Rev Mal Resp 2002;19:727–34.

24. Robertson KD, Jones PA. FASAY. a simple functional assay in yeast for identification of p53 mutation in tumors. Neoplasma 1999,46 80–8.

25. Fulci G, Ishii N, Maurici D, et al. Initiation of human astrocytoma by clonal evolution of cells with progressive loss of p53 functions in a patient with a 28311 TP53 germ-line mutation. evidence for a precursor lesion. Cancer Res 2002,62 2897–905.

26. Inga A, Monti P, Fronza G, Darden T, Resnick MA. p53 mutants exhibiting enhanced transcriptional activation and altered promoter selectivity are revealed using a sensitive, yeast-based functional assay. Oncogene 2001,20:501–13.

27. Bennett WP, Hussain SP, Vahakangas KH, Khan MA, Shields PG, Harris CC. Molecular epidemiology of human cancer risk. gene-environment interactions and p53 mutation spectrum in human lung cancer. J Pathol 1999;187 8–18.

28. Denissenko MF, Pao A, Tang MS, Pfeifer GP. Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. Science (Wash. DC) 1996,274:430–2.

29. Williams C, Norberg T, Ahmadian A, et al. Assessment of sequence-based p53 gene analysis in human breast cancer: messenger RNA in comparison with genomic DNA targets. Clin Chem 1998,44 455–62.

30. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999;354:99–105.

31. Holmila R, Fouquet C, Cadranel J, Zalcman G, Soussi T. Splice mutations in the p53 gene. case report and review of the literature. Hum Mutat 2003;21:101–2.

32. Leung CS, Lung ML. Detection of p53 mutations in Hong Kong colorectal carcinomas by conventional PCR-SSCP analysis versus p53 yeast functional assays. Anticancer Res 1999,19 625–8.

33. Niklinska W, Chyczewski L, Laudanski J, Sawicki B, Niklinski J. Detection of P53 abnormalities in non-small cell lung cancer by yeast functional assay. Folia Histochem Cytobiol 2001,39 147–8.

34. Zochbauer-Muller S, Fong KM, Virmani AK, Geradts J, Gazdar AF, Minna JD. Aberrant promoter methylation of multiple genes in non-small cell lung cancers. Cancer Res 2001,61 249–55.

35. Esteller M, Corn PG, Baylin SB, Herman JG. A gene hypermethylation profile of human cancer. Cancer Res 2001,61 3225–9.

36. Belinsky SA, Nikula KJ, Palmisano WA, et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci USA 1998,95 11891–6.

Exhibit 10:   Favis et al., "Harmonized Microarray/Mutation Scanning Analysis of TP53
              Mutations in Undissected Colorectal Tumors," *Human Mutation* 24:63–75 (2004)

METHODS

Harmonized Microarray/Mutation Scanning Analysis of TP53 Mutations in Undissected Colorectal Tumors

Reyna Favis,[1] Jianmin Huang,[1] Norman P. Gerry,[1] Alfred Culliford,[2] Philip Paty,[2] Thierry Soussi,[3] and Francis Barany[1]*

[1]Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, New York; [2]Colorectal Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York; [3]EA3493, Laboratoire de Genotoxicologie des Tumeurs, Institut Curie, Université PM Curie, Paris, France

Communicated by Richard G.H. Cotton

Both the mutational status and the specific mutation of TP53 (p53) have been shown to impact both tumor prognosis and response to therapies. Molecular profiling of solid tumors is confounded by infiltrating wild-type cells, since normal DNA can interfere with detection of mutant sequences. Our objective was to identify TP53 mutations in 138 stage I–IV colorectal adenocarcinomas and liver metastases without first enriching for tumor cells by microdissection. To achieve this, we developed a harmonized protocol involving multiplex polymerase chain reaction/ligase detection reaction (PCR/LDR) with Universal DNA microarray analysis and endonuclease V/ligase mutation scanning. Sequences were verified using dideoxy sequencing. The harmonized protocol detected all 66 mutations. Dideoxy sequencing detected 41 out of 66 mutations (62%) using automated reading, and 59 out of 66 mutations (89%) with manual reading. Data analysis comparing colon cancer entries in the TP53 database (http://p53.curie.fr) with the results reported in this study showed that distribution of mutations and the mutational events were comparable. Hum Mutat 24:63–75, 2004. © 2004 Wiley-Liss, Inc.

KEY WORDS: zip-code addressing; TP53; p53; mutation detection; microarray; endonuclease V; thermostable ligase; mismatch recognition

INTRODUCTION

Loss of TP53 (MIM# 191170) is observed in approximately one-half of all human cancers, making it the most commonly inactivated tumor suppressor gene [Soussi, 2003; Soussi and Béroud, 2001]. By disrupting TP53 function, cellular stress signals such as DNA damage, oxidative stress, hypoxia, and nucleotide depletion [Vogelstein et al., 2000; Vousden and Lu, 2002] go unheeded, creating a permissive environment for sequence errors that lead to oncogenic mutation. Disruption of TP53 activity breaches a second line of defense, as TP53 also responds to unregulated growth signals caused by the overexpression of certain oncogenes [Vousden, 2002]; thus, TP53 disruption can also contribute to uncontrolled proliferation of cells harboring activated oncogenes. Normally, TP53 will respond to such signals either by arresting the cell cycle to permit DNA repair in mildly damaged genomes or by inducing apoptosis to eliminate cells with severely damaged genomes. In addition to preventing propagation of genomic errors, TP53 is also implicated in the regulation of genes that inhibit angiogenesis and metastatic disease

progression. The tumorigenic potential of a cell is greatly influenced by the functional status of TP53.

Many prospective cancer therapy studies indicate that the functional status of TP53 will also influence a tumor's response to therapy. For example, the commonly used cancer drug, 5′-fluorouracil, is ineffective in TP53-deficient human cells; however, the DNA damaging agent, adriamycin, induces apoptosis irrespective of the TP53 status [Bunz et al., 1999]. Similarly, cells with transcriptionally inactive forms of TP53 are less sensitive to vinca alkaloids, but become more sensitive to

Optimize each assay to detect mutations that the other
misses to achieve high sensitivity and high specificity

paclitaxel [Zhang et al., 1998]. The presence of wild-type TP53 activity is essential for therapies dependent on the anti-angiogenic agent, TNP-470 [Zhang et al., 2000] and anti-angiogenic combination therapy [Yu et al., 2002], while the "ONYX-015" adeno-like virus depends on the complete absence of wild-type TP53 activity in order for transduction to occur. Recent studies have indicated that breast cancer patients harboring TP53 mutations have significantly worse prognoses than those with wild-type status [Borresen-Dale, 2003]. Thus the functional status of TP53 may influence treatment outcome.

While the functional status of TP53 in a cell affects response to therapies and tumor prognosis, several studies have suggested that not all TP53 mutations are alike. TP53 mutant proteins with a flexible conformation correlate with poor prognosis, and different missense mutations are known to have differing effects on the conformational stability of TP53 [Chen et al., 2001]. It has also been demonstrated that cell lines ectopically expressing various mutant forms of TP53 are insensitive to different chemotherapeutic agents [Blandino et al., 1999]. In addition, different TP53 mutations have different cellular consequences. Finally, small synthetic molecules have been generated that are capable of restoring wild-type TP53 function both in vitro and in vivo [Bullock and Fersht, 2001; Bykov et al., 2002]. Thus, the specific TP53 missense mutation in a cell impacts response to therapies and tumor prognosis.

These initial studies indicate that knowledge of the specific TP53 mutation is of growing importance. Although current methods for mutation detection possess some very desirable characteristics (e.g., denaturing gradient gel electrophoresis [DDGE], DHPLC, SSCP, dideoxy-fingerprinting [ddF], restriction endonuclease fingerprinting [REF]), these methods are of limited utility in large-scale prospective clinical trials involving solid tumors [Elsaleh et al., 2001; Kimler et al., 2000; Nabholtz et al., 1999, 2002; Soong et al., 2000] due to low throughput and/or sensitivity (see Kirk et al. [2002] for review). Immunohistochemical analysis of 142 colorectal tumors demonstrated that only 51% of tumors that significantly overexpressed the TP53 protein contained DNA mutations [Kaserer et al., 2000]. Likewise, 32% of tumors that contained a mutated TP53 gene did not concordantly overexpress the TP53 protein. Direct sequencing and gene hybridization chips fail to identify mutations in TP53 over 20% of the time, due in part to dilution of mutant alleles by infiltrating stromal cells in solid tumors [Ahrendt et al., 1999]. Thus, the functional status of TP53 does not necessarily correlate with immunostaining, sequencing, or hybridization chip results [Kaserer et al., 2000].

For effective drug therapy of solid tumors, there is an urgent need to accurately assess TP53 functional status and to precisely determine the nature of the TP53 mutation. In order to substantiate that certain factors are of major effect in influencing outcome, it is necessary to establish statistical significance by surveying a large number of tumors. In this study, we sought to improve both the accuracy and the throughput of TP53 mutation detection by developing a harmonized protocol that combines the strengths of two sensitive enzymatic assays. Rapid analysis is promoted by creating two complementary, parallel tracts with facility for efficient throughput. Endonuclease V (EndoV)/ligase mutation scanning can detect unknown mutations and allows sample pooling [Huang et al., 2002]. This method has been shown to detect substitutions, insertion/deletion mutations varying in size from one to three bases, and scanning ability in amplicons up to 1.7 kb [Huang et al., 2002]. The polymerase chain reaction/ligase detection reaction (PCR/LDR) [Khanna et al., 1999] has substantial multiplexing capability for predetermined mutations, which is extended further by coupling analysis to a Universal DNA microarray [Favis et al., 2000; Gerry et al., 1999]. Both enzymatic assays have sufficient sensitivity to allow analysis of undissected solid tumors, which substantially improves throughput. Figure 1 illustrates how this harmonized protocol functions (Fig. 1A) and provides a flow chart of how the two parallel tracts advance (Fig. 1B). The current study provides the first report of the TP53 Universal DNA microarray and details the application of an assay that combines the utility of two sensitive enzyme systems to analyze mutations in undissected solid tumors.

MATERIALS AND METHODS
Tumor Procurement and DNA Extraction

Control DNA samples with known TP53 mutations were obtained from preexisting samples archived in T. Soussi's collection. All patients recruited from Memorial Sloan Kettering Cancer Center underwent surgical resection for primary adenocarcinoma of the colon. Written informed consent was obtained from each subject. The majority (95%) of patients were identified as Caucasian, while 5% were identified as non-Caucasian. A total of 120 primary colon tumors (15 Stage I, 22 Stage II, 41 Stage III, and 42 Stage IV) were collected at the time of surgical resection in accordance with Institutional Review Board approved protocols. Two to four viable portions of the tumor were harvested by sharp

---

FIGURE 1. Mutation detection using PCR/LDR and EndoV mutation scanning assays. A: Schematic explaining PCR/LDR and EndoV/Ligase mutation scanning assays. B: Flow chart of combined PCR/LDR and EndoV mutation scanning analysis of TP53 mutation status. DNA extracted from undissected primary tumors is subjected to multiplex PCR amplification of exons 4-8. PCR/LDR analysis (right side) subsequently uses multiplex LDR to simultaneously query for the presence of 110 mutations. The ligation products are analyzed on a universal DNA microarray and samples containing TP53 mutations (indicated by the white arrow) are confirmed using uniplex LDR. The uniplex LDR reaction targets only the specific candidate mutation and the control reaction for the amplicon of interest. EndoV mutation scanning (left side of figure) draws from the same multiplex PCR reaction and specific amplicons are selectively amplified. Amplicons for the same exon that are derived from different tumor samples are pooled in groups of three and subjected to EndoV mutations scanning. If mutations are found (indicated by the white arrows), the amplicon harboring the mutation is reanalyzed in each individual sample. Automated sequencing identifies the specific mutation.

dissection and snap frozen in liquid nitrogen within 15 min of removal. From each tumor, four core samples containing a mean normal cell count between 30–50% were taken from different regions, combined, and DNA was extracted using the QIAamp Tissue Kit (Qiagen, Chatsworth, CA) according to manufacturer's guidelines.

Multiplex PCR Amplification

DNA extracted from undissected primary tumors was subjected to multiplex PCR amplification of exons 5, 6, 7, and 8 for gel-based assays, while exon 4 was added to the multiplex for microarray assays in order to include the codon 72 polymorphism in the analysis of exons 4–8. PCR was performed in a 50 μl reaction using 100 ng DNA, 100 μM dNTP, 1 × PCR Buffer II (Applied Biosystems, Foster City, CA) supplemented with 1.5 mM final concentration of $MgCl_2$, 2.5 units AmpliTaq Gold (Applied Biosystems), and 0.4 μM of each primer. Primer sequences, in 5' to 3' orientation, were as follows: exon 4 forward = CCGGAC-GATATTGAACAATGGTTC; exon 4 reverse = GCAA-GAAGCCCAGACGGAAAC, exon 5 forward = CTGTTCACTT GTGCCCTGACTTTC, exon 5 reverse = CCAGCTGCTCAC-CATCGCTATC, exon 6 forward = CCTCTGATTCCTCACT-GATTGCTCTTA, exon 6 reverse = GGCCACTGACAACCA CCCTTAAC, exon 7 forward = GCCTCATCTTGGGCCTGTGT TATC, exon 7 reverse = GTGGATGGGTAGTAGTATGGAA-GAAATC, exon 8 forward = GGACAGGTAGGACCTGATTT CCTTAC, exon 8 reverse = CGGTTCTTGTCCTGCTTGCT-TAC. PCR was performed by heating the reaction for 10 min at 95°C, followed by 35 cycles of 94°C for 30 sec, 60°C for 30 sec, and 72°C for 1 min. Amplification was verified by examining the products on a 3% agarose gel. Taq Polymerase was eliminated by incubating the reaction for 10 min at 70°C with 50 μg/ml Qiagen proteinase K. This treatment was followed by incubation at 90°C for 15 min to denature proteinase K. The TP53 sequence used was GenBank Accession X54156.1 Version X54156.1 GI:35213.

Mutation Detection and Analysis

Using the UMD (Universal Mutation Database) software described by Béroud and Soussi [2003], we analyzed the frequency of TP53 mutational events in colon cancer. Among the 1,427 TP53 mutations that are described for colon cancer, there are 375 different variant classes, ranging from a high frequency of occurrence (such as g.13203G > A [p.R175H] found 177 times) to 230 variants that are found only one time. From this information, it was possible to devise 58 ligase detection reactions (LDR) that allow the detection of 70% of TP53 mutation in colon cancer. Oligonucleotide design and synthesis, ligase detection reaction (LDR), and Tth DNA ligase production were performed as previously described [Barany and Gelfand, 1991; Khanna et al., 1999]. LDR primers were divided into two tubes, based on whether the ligation was directed against the upper or lower strand of DNA. LDR primers designed for gel assays are published elsewhere [Dong et al., 2001], while those for array experiments are available in Supplementary Table S1 (available online at http://www.interscience.wiley.com/jpages/1059-7794/suppmat).

Following a multiplex PCR amplification of the regions of interest, each mutation is simultaneously detected using a thermostable ligase that joins pairs of adjacent oligonucleotides complementary to the sequences of interest. Ligation occurs only when there is perfect complementarity at the junction between the paired oligonucleotides. Ligation products are distinguished based on differential labeling and migration on a polyacrylamide gel, or hybridization to specific addresses on the universal DNA microarray. The reaction was performed as previously described [Favis et al., 2000], except 3 μl of PCR reaction was used. Quality control for LDR was performed using a synthetic template for each mutation to test the ability of the full mix of upper or lower ligation primers to produce the expected, specific signal on the

DNA microarray. In addition, 100 DNA samples derived from various types of cancers and known to contain TP53 mutations targeted in the PCR/LDR assay were analyzed in a blinded fashion.

EndoV/ligase mismatch scanning was performed as described elsewhere [Huang et al., 2002]. Briefly, heteroduplexed variant and wild-type PCR amplicons from tumor DNA are generated using 6FAM and TET labeled primers. Thermotoga maritima (Tma) EndoV recognizes and primarily cleaves heteroduplex DNA one base 3' to a mismatch. Since matched DNA is also nicked at low levels, a highly specific thermostable DNA ligase is used to reseal just those nicks. This lowers background signal and improves the signal-to-noise ratio. Fragment mobility of cleavage products on a DNA sequencing gel reveals the approximate position of the mutation. Amplicons from different tumors corresponding to specific exons were pooled in groups of three. Mutation scanning was performed, and samples identified as containing TP53 mutations were analyzed individually to confirm the mutation and then sequenced. Amplicons for exons 5, 6, 7, and 8 were generated using the PCR primers described for PCR/LDR above and were sequenced using the dRhodamine Terminator Cycle DNA Sequencing kit (Applied Biosystems, Foster City, CA) according to manufacturer's guidelines.

Mutations in the TP53 gene were analyzed with the UMD software [Béroud and Soussi, 2003]. The TP53 database version used for this analysis (December 2002) contains 14,968 mutations, including 1,516 for colon cancer. Genetic alterations reported herein use nucleotide numbering identical to Accession X54156.1.

Microarray Fabrication and Hybridization

Fabrication and hybridization were performed as previously described [Favis et al., 2000; Gerry et al., 1999], except hybridization was carried out in the presence of 100 μg/ml sheared salmon sperm DNA. Arrays were reused three times and were stripped between uses by submerging for 1 min in a solution of boiling 100 mM Bicine/0.1% SDS. Stripped arrays were rinsed in nanopure water, excess water was removed using forced air, and the arrays were stored in slide boxes at room temperature.

Quality control for array fabrication was performed on representative arrays by staining with SYBR Green II (Molecular Probes, Eugene, OR) to determine whether all 64 zip-code addresses had spotted. To verify that no cross-contamination of addresses occurred during spotting, selected arrays were subjected to four hybridizations (stripping between hybridizations) using 6FAM-labeled complementary zip-codes. Arrays were hybridized such that odd rows, even rows, odd columns, or even columns were selectively targeted to produce specific signals without extraneous signals.

RESULTS

Validation of the PCR/LDR Tract: 58 Different TP53 Mutations Can Be Detected Using a Gel-Based Assay

The TP53 gene is mutated in hundreds of positions, and thus represents an enormous challenge to mutation detection strategies [Soussi and Béroud, 2001]. In addition, different cancers exhibit different preferential sites of mutation in this gene [Hussain et al., 2000]. To circumvent problems associated with molecular heterogeneity in TP53, we initially developed a gel-based assay to detect high frequency mutations in colon cancer and established validation criteria for the assay. (Validation of the EndoV tract can be found elsewhere [Huang et al., 2002].) By focusing on a single cancer and using a bioinformatics approach to dictate the assay design, it

was possible to engineer the assay sensitivity such that a significant percentage (approximately 70%) of database-predicted mutations for colon cancer could be detected [Iacopetta, 2003].

LDR is a versatile method for discriminating single-base mutations or polymorphisms and is ideal for multiplexing, since several primer sets can ligate along a gene without the interference encountered in polymerase-based systems (see Fig. 1 for description). Additionally, LDR readily discriminates between wild-type and frameshift or point mutation sequences [Favis et al., 2000; Gerry et al., 1999; Khanna et al., 1999, Zirvi et al., 1999].

A total of 111 LDR primers were designed for this initial assay and the reaction was optimized to achieve multiplex PCR amplification of exons encompassing the DNA binding domain (exons 5, 6, 7, and 8), followed by the multiplex LDR detection of 58 mutations and four amplicon controls (see Dong et al. [2001] for the list of targeted mutations and primer sequences). To validate the assay's ability to detect mutations in clinical samples, 100 blinded tumor DNA samples containing TP53 mutations known to be targeted by the assay were analyzed (see Fig. 2B). No information regarding the origin or the nature of the samples was provided until after results were submitted and samples were unblinded, hence all samples were treated identically and subjected to multiplex PCR followed by multiplex LDR. The tumor DNA was derived from a variety of cancers, from both fresh frozen and paraffin-embedded tissue and included both normal and mutant samples. In addition, artificial mixes of tumor genomic DNA diluted 1:20 (p.R273H (g.14487G > A), double mutant p.R196X + p.R248Q ([g.13346C > T] + [g.14070G > A]), and p.R175H (g.13203G > A)) and 1:100 (p.R175H (g.13203G > A) and p.R273C (g.14486C > T)) in normal genomic DNA were included. The results demonstrated that the PCR/LDR TP53 assay could detect all mutations that were represented by LDR primer sets, even when diluted at 1:100 (Fig. 2C). (In natural tumor samples where it was found that 5% of cells were tumorigenic, we have demonstrated successful mutation detection using PCR/LDR [Fouquet et al., 2003].) The two mutant samples that were refractory contained mutations that were not included in the LDR primer sets.

Although PCR error was not expected to affect assay sensitivity, this was verified by comparing the EndoV mutation scanning results using the proofreading polymerases. Different polymerases for PCR achieve different rates of fidelity, with Taq polymerase error rates of 1.3 × 10^{-5} to 3 × 10^{-6} per base and proofreading polymerases on the order of 5 × 10^{-7}. These error rates are very small compared to the target sensitivity of 1 in 100, and were not expected to create a signal on the sequencing gel comparable to a true mismatch, even at a dilution of 1:100. As confirmation, we experimentally compared Taq polymerase with commercially available high-fidelity mixes containing Pfu or Tgo polymerases, with no difference in background observed (data not shown).

Validation of the PCR/LDR Tract: PCR/LDR Coupled to Analysis on a Universal DNA Microarray to Accommodate Analysis of 110 Different TP53 Mutations

To be of clinical utility, it is necessary to be able to survey a large number of mutations. In deference to the requirement that LDR products have precisely-defined mobility, analysis using a gel-based system places a practical limit of approximately 60–80 alleles per lane of a sequencing gel. To overcome these limitations, analysis of the LDR products was transferred to a universal microarray system. We expanded the number of LDR primers for universal microarray analysis so that the reaction could in toto accommodate 70% of all mutations found in the TP53 database that were associated with colon cancer (exon 4 was added to PCR multiplex to accommodate codon 72 polymorphisms), 65% associated with osteosarcomas (14 mutations and two polymorphisms added with 30 new primers), and 80% associated with lung cancer (38 mutations added using 71 primers). The mutations detected and the universal microarray design are shown in Supplementary Figure S1 (available online at http://www.interscience.wiley.com/jpages/1059-7794/suppmat) while the LDR primer sequences are shown in Supplementary Table S1. The universal array (Fig. 3A) was validated as described for the gel-based assay and then tested for the ability to detect mutations in undissected tumor samples (Fig. 3B). Each mutation is uniquely identified based on the color of the fluorescent signal, the address emitting the signal, and whether the signal was observed for the "upper" or "lower" strand reaction.

Dideoxy Sequencing Is Insufficient for Mutation Analysis in Undissected Colon Tumors

DNA was extracted from 120 stage I–IV colorectal tumors and was examined for TP53 mutations and the status of amino acid 72. The TP53 gene was analyzed using the harmonized protocol and DNA sequencing. As described above, throughput for the EndoV tract was facilitated by pooling amplicons from specific exons in groups of three, these were scanned, and those bearing mutations were rescanned individually and sequenced (Fig. 1).

Table 1 shows a comparison between the harmonized protocol and dideoxy-sequencing. The harmonized protocol identified all 66 TP53 mutations found in the tumor set. We can achieve this high level of sensitivity and specificity because the two tracts of the harmonized protocol complement each other. Of the 66 mutations, 14 were mutations not represented in the LDR primer set, and two mutations fell below the limits of detection for this approach (p.H193R (g.13338A > G) and p.G245S (g.14060G > A) were detectable using synthetic template controls), however EndoV mutation scanning succeeded in detecting these mutations. Similarly, EndoV mutation scanning was unable to detect 20 of the mutations in four specific GC-rich sites known to be refractory to cleavage by this enzyme [Huang et al.,

EXON	MUTATION	LANE	SIZE	EXON	MUTATION	LANE	SIZE
7	M237I G-A/T	U	56				
8	V272M, V272L	U	54				
9	G244C, G244D, G244V	U	52	6, 7	Y205S, Y205C	L	53
				8	R306X	L	51
6	WT	U	49	7, 5	S241F, S241C, K132R	L	49
5	C136Y, C135F	U	46				
5	C176Y, C176F	U	44	8	R282W	L	46
5	V173M, V173L, V143M, V143A	U	42	6	R196X	L	43
5	G245D	U	40	7	I254V G154S	L	41
				5	R175C, R175L, H178Y	L	39
6	WT	U	37	7	R249W R249Q	L	37
				5	H179R, H179L	L	35
				5	WT	L	33

(A)

(B)

(C)
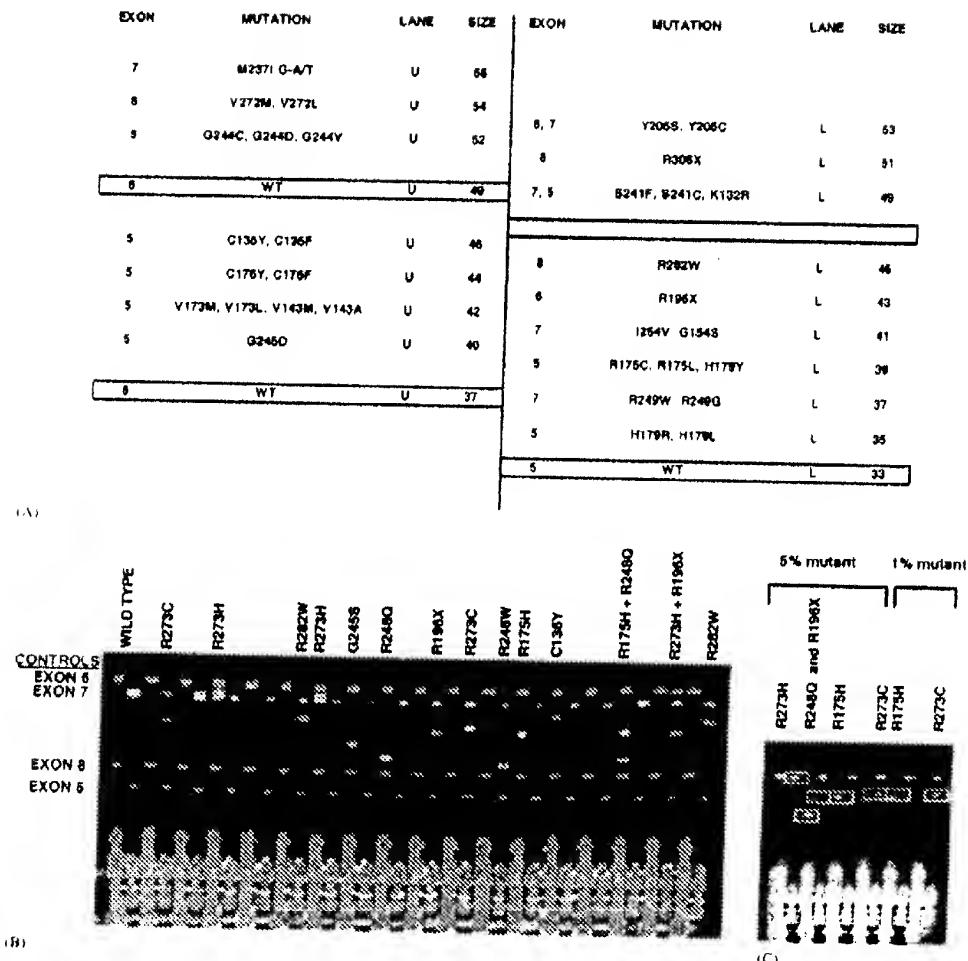
FIGURE 2. Gel-based assay can simultaneously query for 58 different TP53 mutations. A: Diagram of gel results for PCR/LDR assay. The gel-based assay can simultaneously query for 58 different TP53 mutations. LDR reactions rely on ligation of primers identical to either the upper strand (U) or lower strand (L) of the DNA sequence. U and L reactions are performed in separate tubes and loaded into separate lanes. The columns to the left of the bold red line represent the two lanes of a 10% sequencing gel. The boxed products are the amplicon controls for TP53 exons 5, 6, 7, and 8. The TP53 exon containing the mutation and the size of each product (bases) is shown in the first and last columns of each lane, respectively. B: Gel results of a mutation detection experiment using 100 blinded samples of DNA isolated from tumors. Amplicon controls for each exon are designated on the left of the image (identical to boxed regions in A). Mutation status of the sample is indicated over the second of the two lanes dedicated to each sample. For both panels, blue products are FAM-labeled reactions, while green products are reactions labeled with TET. Mutations detected are represented by amino acid and position. C: Detection of mutant DNA diluted 1:20 and 1:100 in normal DNA. DNA samples with known TP53 mutations were diluted into wild-type DNA and provided in a blinded manner for TP53 mutation detection. The samples were subsequently PCR amplified and subjected to PCR/LDR. The mutations found (listed above the corresponding two gel lanes for each sample) were confirmed to match the known mutation in the diluted DNA. LDR products for mutations are indicated with boxes. Dilution appears at the top of the figure over corresponding gel lanes.

2002], but PCR/LDR readily detected these mutations. Because PCR/LDR/Universal array can find mutations that are refractory to cleavage by EndoV mutation scanning, while the latter finds deletions or uncommon mutations not covered by our LDR primers, this harmonized protocol can find all mutations present in

(A)

(B)
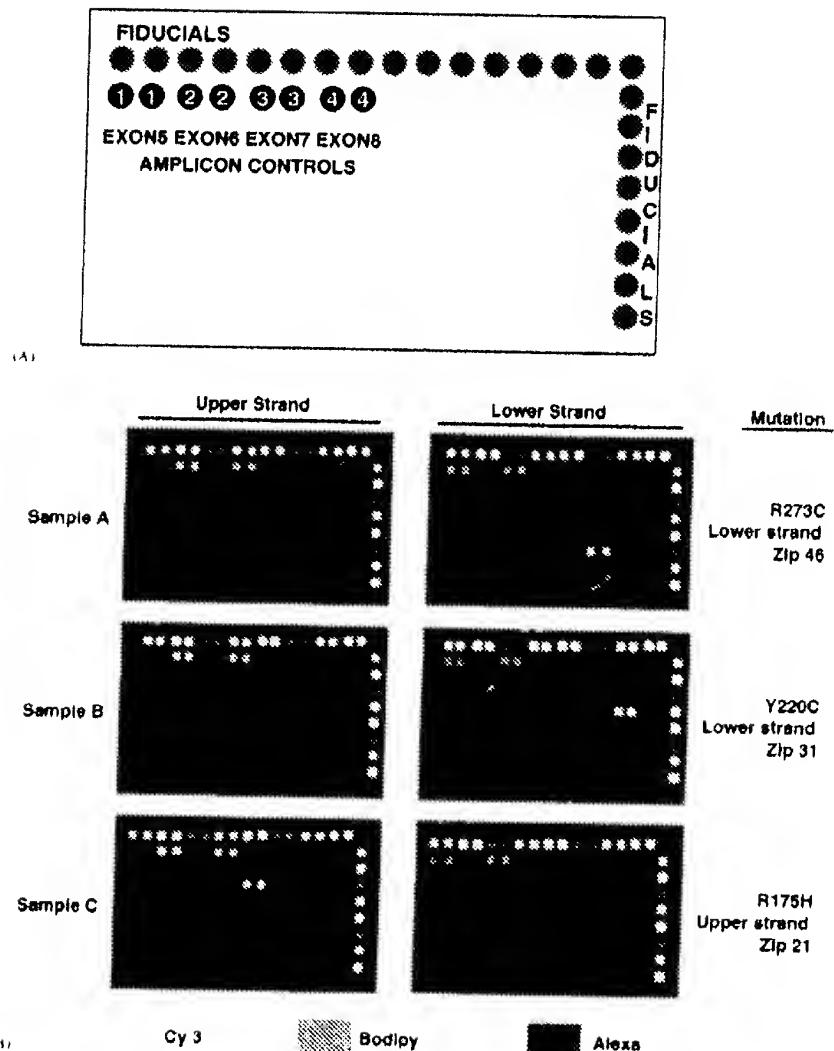
FIGURE 3. Universal DNA microarray analysis can detect 110 differentTP53 mutations. A: Diagram of chip layout for theTP53 assay. The assay uses two 64-address arrays, but for simplicity, a composite diagram is presented. Addresses are double-spotted on io a three-dimensional surface comprised of a loosely cross-linked polymer of acrylamide and acrylic acid. The three-dimensional surface combined with the zip-code system allows hybridized arrays to be stripped of target and reused. Fiducials labeled with fluorochromes Cy3, Bodipy, and Alexa are spotted along the top and tha right side of the array to provide orientation. Amplicon controls for exons 5, 6, 7, and 8 are directed to zip-codes in the first row of the array and labeled with both Cy3 and Bodipy. The zip-code addresses occupied by these controls are indicated by the numbers in the circles: upper strand control reactions hybridize to zip-code 2 (exon 6) and zip-code 4 (exon 8) (later versions of the array include zip-code 5 (exon 4)); lower strand control reactions hybridize to zip-code 1 (exon 5) and zip-code 3 (exon 7). A maximum of three different products can be directed to each address. Cy3 signal is represented in green, Bodipy-labeled products are red, and Alexa signals are blue. B: Results of PCR/LDR/Universal DNA microarray analysis of undissected colon tumor DNA. Three different samples are depicted. LDR reactions directed against the upper (left side) or lower strand (right side) of DNA are performed in separate tubes and hybridized to different arrays. Mutations found in each sample are named on the far right of the panel and are identified by the location of the zip-code exhibiting the signal, the color of the fluorochrome associated with the signal, and whether the signal appears on an array for upper or lower strand reactions.

TABLE 1. Comparison Between the Harmonized Protocol and Dideoxy-sequencing

Sample #: (138 total)	Combined Assay	DNA sequencing, Automated read	Resequencing of both strands, Manual read	TP53 Mutation
1	(7) Y234C (W)*	ND	ND	g.14028A>G
2	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
4	(8) R273C (REV)*	(8) R273C	NN	g.14486C>T
6	(7) R248Q*'	ND	(7) R248Q	g.14070G>A
7	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
9	(7) R248Q*'	ND	(7) R248Q	g.14070G>A
13	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
15	(6) R213X (W)*'	(6) R213X	NN	g.13397C>T
17	(7) Y234C (W)*	ND	ND	g.14028A>G
18	(8) R282W*'	(8) R282W	NN	g.14513C>T
20	(8) R273C (REV)*	(8) R273C	NN	g.14486C>T
24	(7) R248W*'	ND	(7) R248W	g.14069C>T
26	(8) R273H*'	(8) R273H	NN	g.14487G>A
27	(7) R248Q*'	ND	(7) R248Q	g.14070G>A
29	(8) R273C (REV)*	(8) R273C	NN	g.14486C>T
31	(8) R273H*'	(8) R273H	NN	g.14487G>A
32	(5) E171X*'	(5) E171X	NN	g.13190G>T
33	(5) G154G (LF, S)'	(5) G154G (S)	NN	g.13141C>A
35	(5) G154D (LF)'	(5) G154D	NN	g.13140G>A
36	(5) Y163C (LF)'	(5) Y163C	NN	g.13167A>G
38	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
40	(8) R306X (REV)*	ND	(8) R306X	g.14585C>T
41	(6) E224E (LF, S)'	ND	(6) E224E (S)	g.13432G>A
42	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
43	(6) H193R'	(6) H193R	NN	g.13214C>T
44	(5) H179Y*'	(5) H179Y	NN	g.13338A>G
45	(7) R248Q*'	ND	(7) R248Q	g.14070G>A
46	(7) R248Q*'	ND	(7) R248Q	g.14070G>A
49	(5) R175H (W, REV)*	(5) R175H	NN	g.13203G>A
53	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
55	(8) E285K*'	(8) E285K	NN	g.14522G>A
59	(8) R273C (REV)*	(8) R273C	NN	g.14486C>T
60	(8) R282W,*	ND	(5) K164X	g.14513C>T
	(8) R306X (REV)*		(8) R306X	g.14585C>T
	(5) K164X'			g.13169A>T
65	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
66	(7) R248Q*'	ND	(7) R248Q	g.14070G>A
67	(8) R282W*'	ND	(8) R282W	g.14513C>T
68	(6) Y205F (LF)'	ND	(6) Y205F**	g.13374A>T
71	(8) R273C (REV)*	(8) R273C	NN	g.14486C>T
73	(5) Q167 insA'		(5) Q167 insA'	g.13178.13179insA
	(8) R282W*'	(8) R282W	(8) R282W	g.14513C>T
77	(7) S261R (LF)'	(7) S261R	NN	g.14108A>C
78	(7) R248W*'	ND	(7) R248W	g.14069C>T
79	(7) S240 delA (LF)'	ND	(7) S240 delA**	g.14045delA
80	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
81	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
84	(7) S261R (LF)'	(7) S261R	NN	g.14108A>C
89	(6) Q192X (LF)'	(6) Q192X	NN	g.13334C>T
90	(6) H214 insA (LF)'	ND	(6) H214 insA**	g.13401.13402insA
91	(7) G245S'	(7) G245S	NN	g.14060G>A
93	(5) R175H (REV)*	(5) R175H	NN	g.13203G>A
94	(6) H214 insA (LF)'	ND	(6) H214 insA**	g.13401.13402insA
96	(8) R273H*'	(8) R273H	NN	g.14487G>A
97	(8) R267G (LF)'	(8) R267G	NN	g.14468C>G
98	(8) R282W*'	(8) R282W	NN	g.14513C>T
102	(7) R248Q*'	ND	(7) R248Q	g.14070G>A
103	(7) R248Q*'	ND	(7) R248Q	g.14070G>A
114	(6) R213R (LF, S)'	(6) R213R (S)	NN	g.13399A>G
116	(6) P190L (LF)'	(6) P190L	NN	g.13329C>T
117	(6) R213R (LF, S)'	(6) R213R (S)	NN	g.13399A>G
122	(7) R248W*'	ND	(7) R248W	g.14069C>T
123	(7) E258D (LF)'	(7) E258D	NN	g.14102A>T
124	(8) R273C (REV)*	(8) R273C	NN	g.14486C>T
126	(6) E224E (LF, S)'	ND	(6) E224E (S)	g.13432G>A
137	(8) R273C (REV)*	(8) R273C	NN	g.14486C>T
(75 samples)	ND	Not determined #	Not determined #	
p53 Mutants	66/66	41/66	59/66	
% correct	Set as standard	62%	89%	

(5), (6), (7), (8) = exon 5, 6, 7, or 8; * = detected by PCR/LDR; ' = Detected by Endo V; ** = Required gel purification of PCR product to obtain sequencing result. # = Sequencing of 5 random samples of 75 called negative by harmonized protocol reveal no new mutations; ND = Not detected; LF = Current PCR/LDR universal array primer set does not contain primers to detect this low frequency mutation, NN = Not necessary since mutation was identified using automated DNA sequencing; REV = Refractory to Endo V cutting, (S) = Silent mutation; and (W) = Weak signal. All nucleotide numeric designations use numbering identical to. GenBank Accession X54156.1.

this cohort of undissected tumor samples. By comparison, direct sequencing with automated reading of the sequence found only 41 out of 66 mutations (62% sensitivity) and consistently failed to identify mutations in codon R248, which is the most frequently mutated TP53 codon for all cancers. When direct sequencing with automated reading was supplemented with resequencing of both strands and manual reading, the score improved to 59 out of 66 mutations identified (89% sensitivity). However, for five of the samples, the mutation could only be detected by gel purifying the PCR product prior to sequencing. In addition, prior knowledge of the site of mutation was required when manually reading sequencing results to identify certain mutations that were present at low levels.

Analysis of 138 Colorectal Tumors and Liver Metastases Using the Harmonized Protocol

A summary of the colorectal tumor data can be found in Table 2. We found that 44% of the samples were mutant for TP53. The frequency of mutation for TP53 was in agreement with previous findings for colorectal cancer [Iacopetta, 2003; Soussi et al., 2000]. For TP53, 60 samples had point mutations, four samples exhibited deletions, five samples contained polymorphisms, one sample was double mutant (insertion of A in codon Q167 + p.R282W ([g.13178_13179A] + [g.14513C > T])), and one sample was identified as a triple mutant (p.K164X + p.R282W + p.R306X ([g.13169A > T] + [g.14513C > T] + [g.14585C > T])).

As a quality control step before data release, the first pass results from the present study were compared to the 1,516 mutations found in colon cancer that were stored in the database. The distribution of the mutations along the TP53 gene and the pattern of mutation events were quite similar (see Supplementary Figure S2, available online at http://www.interscience.wiley.com/jpages/1059-7794/suppmat), reinforcing the accuracy of the harmonized protocol. In addition, by performing detailed comparisons of the current results with the TP53 database, we were able to identify and quickly eliminate a spurious weak LDR signal corresponding to p.R273S (g.14486C > A). This mutation had never been detected in the 1,620 entries for colorectal cancer and only 11 times in the database overall. When preliminary PCR/LDR results suggested the low-level presence of this mutation in seven samples, the finding was considered highly improbable; the results from the EndoV tract were closely inspected and DNA sequencing in both directions was pursued. Failure to confirm the borderline LDR finding led us to carefully scrutinize the p.R273S (g.14486C > A) reaction. It was revealed that human error led to the synthesis of LDR oligonucleotides that did not correspond to the correct signal. Since our synthetic template to test ligation reactions was designed as reverse complements of the joined LDR oligonucleotides, early tests of the chip did not identify this reaction as problematic. As a result of this finding, we will in the future design synthetic templates de novo, without

referring to the LDR product for the reaction. Thus, by relying on a bioinformatics approach, spurious signals can be readily distinguished from authentic mutations and processes can be continuously refined and improved. In contrast, mutation detection at the very same R273 codon by gene chip hybridization was unable to distinguish a weak signal resulting from false hybridization of wild-type sequence [Wen et al., 2000] from the presence of low-level mutant allele.

A common polymorphism in TP53 is located at codon 72. Since LDR had been shown to be sufficiently sensitive to detect the presence of a variant sequence when diluted 1:100 in wild-type sequences (see description of gel-based assay above), determination of the status of both codon 72 alleles was possible, even in the event of 17p loss of heterozygosity (LOH) in the tumor cells, due to the 30 to 50% estimated stromal contamination. In agreement with previous studies that showed the allele frequencies in Caucasians to be approximately 70% arginine allele and 30% proline allele [Beckman et al., 1994; Harris et al., 1986], we found allele frequencies of 75 and 25% for the R and P alleles, respectively.

DISCUSSION

We have designed a harmonized protocol that relies on two sensitive enzymatic reactions to detect the presence of mutations in undissected solid tumors, in which contaminating wild-type stroma can account for the majority of DNA template present in a sample. The universal microarray-based tract uses a thermostable ligase enzyme to detect predetermined mutations by discriminating between wild-type and mutant templates, resulting in the separation of the mutation detection and array hybridization. The mutation scanning tract detects unknown mutations and relies on thermostable EndoV and ligase to produce high sensitivity. The harmonized protocol successfully detected all 66 mutations found in the present study, many of which were missed by DNA sequencing.

Although DNA sequencing using manual reading performed significantly better than automated reading (89 and 62% sensitivity, respectively), identification of mutated bases ultimately relied on prior knowledge of mutation position, as indicated by results from the harmonized protocol. Overall, the fusion of PCR/LDR/ Universal array and EndoV mutation scanning proved to be a rapid means of identifying mutation. While it may be argued that other combinations of mutation detection methods might result in equal sensitivity (e.g., SSCP + sequencing), these applications still fall short. Whereas electrophoretic mobility assays can detect low level mutations, these approaches may miss a significant portion of mutations: SSCP misses 30% of possible mutations [Bjursell et al., 2000; Hayashi, 1991; Korn et al., 1993; Makino et al., 1992; Suzuki et al., 1990]; methods such as CSGE, DGGE, CDGE, and DHPLC, which look for differential electrophoretic migration between homo- and heteroduplexes, have been shown

TABLE 2. Summary of Colorectal Tumor TP53 Mutation Analysis

Variable	Outcome	Frequency	Percent
Tumor Stage	I	15	11
	II	22	16
	III	41	30
	IV	42	30
	Metastases	18	13
	Total	138	100
TP53 Mutation/Polymorphism Summary	WT	75	54
	Mutations	61	44
	Polymorphisms	5	4
	Total	138	100
TP53 Amino Acid 72 Polymorphism Status	P homozygote	10	7
	Heterozygote	55	40
	R homozygote	73	53
	Total	138	100
TP53 Nonsense & Missense Mutations	Missense	51	84
	Nonsense/indel	10	16
	Total	61	100
TP53 Structural Motifs with Mutation	DNA contact	28	46
	L2 loop	18	30
	Beta sandwich	4	7
	H2 helix	6	10
	C-terminus	2	3
	H1 helix	1	2
	L3 loop	2	3
	Total	61	100
TP53 Mutations and Polymorphisms*	R175H g.13203 G > A	11	17
	R248Q g.14070 G > A	8	12
	R273C g.14486 C > T	7	11
	R282W g.14513 C > T	5	8
	R248W g.14069 C > T	3	5
	R273H g.14487 G > A	3	5
	R306X g.14585 C > T	2	3
	S261R g.14108 A > C	2	3
	Y234C g.14028 A > G	2	3
	H214 g.13401_13402insA	2	3
	E224E g.13432 G > A	2	3
	R213R g.13399 A > G	2	3
	G154G g.13141 C > A	2	3
	E171X g.13190 G > T	1	2
	E258D g.14102 A > T	1	2
	E285K g.14522 G > A	1	2
	G245S g.14060 G > A	1	2
	G154D g.13140 G > A	1	2
	H179Y g.13214 C > T	1	2
	H193R g.13338 A > G	1	2
	R267G g.14468 C > G	1	2
	P190L g.13329 C > T	1	2
	Q192X g.13334 C > T	1	2
	R213X g.13397 C > T	1	2
	Q167 g.13178_13179insA	1	2
	K164X g.13169 A > T	1	2
	S240 g.14045delA	1	2
	Y163C g.13167 A > G	1	2
	Y205F g.13374 A > T	1	2
	Total	66	100

*All nucleotide numeric designations use numbering identical to GenBank Accession X54156.1

to miss 11% of polymorphisms [Chen and Thilly, 1994; Fodde and Losekoot, 1994; Ganguly, 2002; Ganguly et al., 1993; Guldberg and Guttler, 1994; Khrapko et al., 1994; Kozlowski and Krzyzosiak, 2001; Larsen et al., 2001; Mitchelson, 2001; Ridanpaa and Husgafvel-Pursiainen, 1993; Rozycka et al., 2000]. DNA sequencing, on the other hand, has low sensitivity and will miss low-level mutations. To optimize paired reactions based on these approaches, microdissection would be required. The major strength of the harmonized assay is that mutation detection may be implemented in the absence of microdissection to enrich for tumor DNA.

One advantage of the harmonized protocol is that the two tracts of this combined process use the same multiplex PCR reaction, thus this approach does not consume excessive amounts of limited tumor DNA sample. Also, the two tracts of the process advance in parallel and there is no need for microdissection on either tract, thus this approach is also less time consuming. A current limitation to the wider application of the harmonized assay may be the lack of comprehensive databases for other important tumor suppressor genes and oncogenes. To program the PCR/LDR/Universal array tract in the present study, we used a bioinformatics selection process to guide us to the most significant TP53 mutations in colon cancer. The legitimacy of the programmed mutations is supported by a recent report [Kato et al., 2003], in which 2,314 TP53 mutations were evaluated for functional impact using a yeast-based assay. With the exception of A161 mutations, TP53 mutations included on the universal array showed a loss of activity, demonstrating the efficacy of a bioinformatics selection process. If similar database resources are lacking for other genes of interest, it will first be necessary to build the databases that can clarify which mutations should be targeted for analysis. The rapid advances in sequencing and mutation scanning technology over the past few years will undoubtedly assist in expanding web-based databases of both germline and inherited mutations in cancer-associated genes. Additional bioinformatics resources will likely be launched if it can be shown that tumor profiling can be successfully applied to the clinical decision process.

In comparison to previous analyses using direct hybridization to gene oligonucleotide arrays to detect TP53 mutations in frozen tissue, tumors deficient in neoplastic cells required selective microdissection. Although the intent of the gene hybridization chip was to detect all TP53 mutations, it failed, detecting only 81% [Ahrendt et al., 1999], 84% [Wikman et al., 2000], and 92% [Wen et al., 2000] of TP53 mutations. In all cases, insertion/deletion mutations proved intractable to this detection scheme, and significantly reduced the sensitivity values. Further, gene hybridization arrays required statistical considerations on background to improve specificity from 34 to 86%, but at the cost of reduced sensitivity from 92 down to 84% [Wikman et al., 2000]. In contrast, the LDR primers were predicted to accommodate roughly 70% of colon cancer mutations (as estimated by prevalence in the TP53 database) and PCR/LDR/Universal array succeeded in identifying 68% of TP53 mutations found in the undissected colorectal adenocarcinomas analyzed (Table 1). By creating a harmonized protocol involving both PCR/LDR/Universal array and EndoV mutation scanning, all TP53 mutations in the targeted exons are detected, including insertion/deletion mutations, thus achieving high sensitivity with high specificity. This result demonstrates that the PCR/LDR-bioinformatics approach to universal chip development combined with EndoV mutation scanning presented here out-performs approaches that attempt to target all possible mutations, such as gene chip

hybridization or automated sequencing. The added value of our assay is that time-consuming microdissection is eliminated.

In conclusion, rapid and accurate mutation analysis of tumors is critical to resolve differences in prognosis and response to therapy. Due to the comparatively advanced state of understanding, the TP53 gene is a strategic starting point to demonstrate proof of principle and to advance the translational research to achieve this goal. Approaches that prove successful for TP53 mutation detection (e.g., curation of comprehensive mutation databases, bioinformatics-based approaches to assay design, and development of rapid, sensitive assays) may also be applied to characterizing the mutation status of other tumor suppressor genes and oncogenes. In time, the ability to perform molecular profiling of tumors may facilitate tailoring individualized treatments for individual patients.

ACKNOWLEDGMENTS

REFERENCES

Ahrendt S, Halachmi S, Chow J, Wu L, Halachmi N, Yang S, Wehage S, Jen J, Sidransky D. 1999. Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array. Proc Natl Acad Sci USA 96:7382–7387.

Barany F, Gelfand D. 1991. Cloning, overexpression, and nucleotide sequence of a thermostable DNA ligase gene. Gene 109:1–11.

Beckman G, Birgander R, Sjalander A, Saha N, Holmberg PA, Kivela A, Beckman L. 1994. Is p53 polymorphism maintained by natural selection? Hum Hered 44:266–270.

Béroud C, Soussi T. 2003. The UMD-p53 database: new mutations and analysis tools. Hum Mutat 21:176–181.

Bjursell C, Erlandson A, Nordling M, Nilsson S, Wahlstrom J, Stibler H, Kristiansson B, Martinsson T. 2000. PMM2 mutation spectrum, including 10 novel mutations, in a large CDG type 1A family material with a focus on Scandinavian families. Hum Mutat 16:395–400.

Blandino G, Levine A, Oren M. 1999. Mutant p53 gain of function: differential effects of different p53 mutants on resistance of cultured cells to chemotherapy. Oncogene 18:477–485.

Borresen-Dale A. 2003. TP53 and breast cancer. Hum Mutat 21:292–300.

Bullock AN, Fersht AR. 2001. Rescuing the function of mutant p53. Nat Rev Cancer 1:68–76

Bunz F, Hwang PM, Torrance C, Waldman T, Zhang Y, Dillehay L, Williams J, Lengauer C, Kinzler KW, Vogelstein B. 1999. Disruption of p53 in human cancer cells alters the responses to therapeutic agents. J Clin Invest 104:263–269

Bykov VJ, Issaeva N, Shilov A, Hultcrantz M, Pugacheva E, Chumakov P, Bergman J, Wiman KG, Selivanova G. 2002. Restoration of the tumor suppressor function to mutant p53 by a low-molecular-weight compound. Nat Med 8:282–288.

Chen J, Rosal R, Smith S, Pincus M, Brandt-Rauf P. 2001. Common conformational effects of p53 mutations. J Protein Chem 20:101–105.

Chen J, Thilly WG. 1994. Use of denaturing-gradient gel electrophoresis to study chromium-induced point mutations in human cells. Environ Health Perspect 102(Suppl 3):227–229.

Dong SM, Traverso G, Johnson C, Geng L, Favis R, Boynton K, Hibi K, Goodman SN, D'Allessio M, Paty P, Hamilton S, Sidransky D, Barany F, Levin B, Shuber A, Kinzler K, Vogelstein B, Jen J. 2001. Detecting colorectal cancer in stool with the use of multiple genetic targets. J Natl Cancer Inst 93:858–865.

Elsaleh H, Powell B, McCaul K, Grieu F, Grant R, Joseph D, Iacopetta B. 2001. P53 alteration and microsatellite instability have predictive value for survival benefit from chemotherapy in stage III colorectal carcinoma. Clin Cancer Res 7:1343–1349.

Favis R, Day JP, Gerry NP, Phelan C, Narod S, Barany F. 2000. Universal DNA array detection of small insertions/deletions in BRCA1 and BRCA2. Nat Biotechnol 18:561–564.

Fodde R, Losekoot M. 1994. Mutation detection by denaturing gradient gel electrophoresis (DGGE). Hum Mutat 3:83–94.

Fouquet C, Antoine M, Tisserand P, Favis R, Wislez M, Commo F, Rabhe N, Carette MF, Milleron B, Barany F, Cadranel J, Zalcman G, Soussi T. 2003. Rapid and sensitive p53 alteration analysis in biopsies from a lung cancer patient using a functional assay or a versatile oligonucleotide array: a prospective study. Clin Cancer Res (in press).

Ganguly A, Rock MJ, Prockop DJ. 1993. Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. Proc Natl Acad Sci USA 90:10325–10329.

Ganguly A. 2002. An update on conformation sensitive gel electrophoresis. Hum Mutat 19:334–342.

Gerry N, Witowski N, Day J, Hammer R, Barany G, Barany F. 1999. Universal DNA Microarray method for multiplex detection of low abundance point mutations. J Mol Biol 292:251–262.

Guldberg P, Guttler F. 1994. "Broad-range" DGGE for single-step mutation scanning of entire genes: application to human phenylalanine hydroxylase gene. Nucleic Acids Res 22:880–881.

Harris N, Brill E, Shohat O, Prokocimer M, Wolf D, Arai N, Rotter V. 1986. Molecular basis for heterogeneity of the human p53 protein. Mol Cell Biol 6:4650–4656.

Hayashi K. 1991. PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. PCR Methods Appl 1:34–38.

Huang J, Kirk B, Favis R, Soussi T, Paty P, Zbar B, Cao W, Barany F. 2002. High sensitivity scanning for unknown germline and sporadic cancer mutations using combined cleavage/proofreading by thermostable endonuclease V/DNA ligase. Oncogene 21:1909–1921.

Hussain S, Hollstein M, Harris C. 2000. p53 tumor suppressor gene at the crossroads of molecular carcinogenesis, molecular epidemiology, and human risk assessment. Ann NY Acad Sci 919:79–85.

Iacopetta B. 2003. TP53 mutation in colorectal cancer. Hum Mutat 21:271–276.

Kaserer K, Schmaus J, Bethge U, Migschitz B, Fasching S, Walch A, Herbst F, Teleky B, Wrba F. 2000. Staining patterns of p53 immunohistochemistry and their biological significance in colorectal cancer. J Pathol 190:450–456.

Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, Ishioka C. 2003. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proc Natl Acad Sci USA 100:8424–8429.

Khanna M, Park P, Zirvi M, Cao W, Picon A, Day J, Paty P, Barany F. 1999. Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumors. Oncogene 18:27–38.

Khrapko K, Hanekamp JS, Thilly WG, Belenkii A, Foret F, Karger BL. 1994. Constant denaturant capillary electrophoresis (CDCE): a high resolution approach to mutational analysis. Nucleic Acids Res 22:364–369.

Kimler BF, Fabian CJ, Wallace DD. 2000. Breast cancer chemoprevention trials using the fine-needle aspiration model. J Cell Biochem Suppl 34:7–12.

Kirk BW, Feinsod M, Favis R, Kliman RM, Barany F. 2002. Single nucleotide polymorphism seeking long term association with complex disease. Nucleic Acids Res 30:3295–3311.

Korn SH, Moerkerk PT, de Goeij AF. 1993. K-ras point mutations in routinely processed tissues: non-radioactive screening by single strand conformational polymorphism analysis. J Clin Pathol 46:621–623.

Kozlowski P, Krzyzosiak WJ. 2001. Combined SSCP/duplex analysis by capillary electrophoresis for more efficient mutation detection. Nucleic Acids Res 29:E71.

Larsen LA, Johnson M, Brown C, Christiansen M, Frank-Hansen R, Vuust J, Andersen PS. 2001. Automated mutation screening using dideoxy fingerprinting and capillary array electrophoresis. Hum Mutat 18:451–457.

Makino R, Yazyu H, Kishimoto Y, Sekiya T, Hayashi K. 1992. F-SSCP: fluorescence-based polymerase chain reaction-single-strand conformation polymorphism (PCR-SSCP) analysis. PCR Methods Appl 2:10–13.

Mitchelson KR. 2001. The application of capillary electrophoresis for DNA polymorphism analysis. Methods Mol Biol 162:3–26.

Nabholtz JM, Lindsay MA, Hugh J, Mackey J, Smylie M, Au HJ, Tonkin K, Allen M. 1999. The academic global virtual concept in clinical cancer research and its application to breast cancer. The Breast Cancer International Research Group. Semin Oncol 26(Suppl 8):4–8.

Nabholtz JM, Reese DM, Lindsay MA, Riva A. 2002. HER2-Positive Breast Cancer: update on Breast Cancer International Research Group trials. Clin Breast Cancer 3(Suppl 2):S75–S79.

Ridanpaa M, Husgafvel-Pursiainen K. 1993. Denaturing gradient gel electrophoresis (DGGE) assay for K-ras and N-ras gene detection of K-ras point mutations in human lung tumour DNA. Hum Mol Genet 2:639–644.

Rozycka M, Collins N, Stratton MR, Wooster R. 2000. Rapid detection of DNA sequence variants by conformation-sensitive capillary electrophoresis. Genomics 70:34–40.

Soong R, Powell B, Elsaleh H, Gnanasampanthan G, Smith DR, Goh HS, Joseph D, Iacopetta B. 2000. Prognostic significance of TP53 gene mutation in 995 cases of colorectal carcinoma. Influence of tumour site, stage, adjuvant chemotherapy and type of mutation. Eur J Cancer 36:2053–2060.

Soussi T, Dehouche K, Beroud C. 2000. p53 website and analysis of p53 gene mutations in human cancer: forging a link between epidemiology and carcinogenesis. Hum Mutat 15:105–113.

Soussi T, Béroud C. 2001 Assessing TP53 status in human tumours to evaluate clinical outcome. Nat Rev Cancer 1 233–240

Soussi T 2003 Focus on the p53 gene and cancer advances in TP53 mutation research Hum Mutat 21 173–175.

Suzuki Y, Orita M, Shiraishi M, Hayashi K, Sekiya T 1990. Detection of ras gene mutations in human lung cancers by single-strand conformation polymorphism analysis of polymerase chain reaction products Oncogene 5:1037–1043.

Vogelstein B, Lane D, Levine AJ 2000. Surfing the p53 network. Nature 408:307–310.

Vousden K 2002. Activation of the p53 tumor suppressor protein. Biochim Biophys Acta 1602:47–59.

Vousden K, Lu X. 2002. Live or let die: the cell's response to p53 Nat Rev Cancer 2:594–604

Wen WH, Bernstein L, Lescallett J, Beazer-Barclay Y, Sullivan-Halley J, White M, Press MF. 2000. Comparison of TP53 mutations identified by oligonucleotide microarray and conventional DNA sequence analysis. Cancer Res 60 2716–2722

Wikman FP, Lu ML, Thykjaer T, Olesen SH, Andersen LD, Cordon-Cardo C, Orntoft TF. 2000 Evaluation of the performance of a p53 sequencing microarray chip using 140 previously sequenced bladder tumor samples Clin Chem 46.1555–1561

Yu JL, Rak JW, Coomber BL, Hicklin DJ, Kerbel RS. 2002. Effect of p53 status on tumor response to antiangiogenic therapy. Science 295:1526–1528.

Zhang CC, Yang JM, White E, Murphy M, Levine A, Hait WN. 1998. The role of MAP4 expression in the sensitivity to paclitaxel and resistance to vinca alkaloids in p53 mutant cells. Oncogene 16:1617–1624.

Zhang Y, Griffith EC, Sage J, Jacks T, Liu JO 2000. Cell cycle inhibition by the anti-angiogenic agent TNP-470 is mediated by p53 and p21WAF1/CIP1. Proc Natl Acad Sci USA 97, 6427–6432.

Zirvi M, Nakayama T, Newman G, McCaffrey T, Ostrer H, Paty P, Barany F. 1999 Ligase-based detection of mononucleotide repeat sequences. Nucleic Acids Res 27.e40

Exhibit 11:   Cheng et al., "Multiplexed Profiling of Candidate Genes for CpG Island
               Methylation Status Using a Flexible PCR/LDR/Universal Array Assay," *Genome
               Research* 16(2):282–9 (2006)

Multiplexed profiling of candidate genes for CpG island methylation status using a flexible PCR/LDR/Universal Array assay

Yu-Wei Cheng,[1] Carrie Shawber,[2] Dan Notterman,[3] Philip Paty,[4] and Francis Barany[1,5]

[1]Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, New York 10021, USA; [2]Department of OB/GYN, Columbia University Medical Center, New York, New York 10032, USA, [3]Departments of Pediatrics and Molecular Genetics, University of Medicine and Dentistry of New Jersey (UMDNJ)-Robert Wood Johnson Medical School, New Brunswick, New Jersey 08901, USA, [4]Department of Surgery, Colorectal Surgery Service, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

DNA methylation in CpG islands is associated with transcriptional silencing. Accurate determination of cytosine methylation status in promoter CpG dinucleotides may provide diagnostic and prognostic value for human cancers. We have developed a quantitative PCR/LDR/Universal Array assay that allows parallel evaluation of methylation status of 75 CpG dinucleotides in the promoter regions of 15 tumor suppressor genes (CDKN2B, CDKN2A, CDKN2D, CDKN1A, CDKN1B, TP53, BRCA1, TIMP3, APC, RASSF1, CDH1, MGMT, DAPK1, GSTP1, and RARB). When compared with an independent pyrosequencing method at a single promoter, the two approaches gave good correlation. In a study using 15 promoter regions and seven blinded tumor cell lines, our technology was capable of distinguishing methylation profiles that identified cancer cell lines derived from the same origins. Preliminary studies using 96 colorectal tumor samples and 73 matched normal tissues indicated CpG methylation is a gene-specific and nonrandom event in colon cancer. This new approach is suitable for clinical applications where sample quantity and purity can be limiting factors.

[Supplemental material is available online at www.genome.org.]

Aberrant methylation of CpG dinucleotides in the 5′ regulatory region of genes often results in transcriptional inactivation and has been implicated in aging, heart and neurodegenerative diseases, as well as in the pathogenesis of various types of cancers (Feinberg and Vogelstein 1983; Gardiner-Garden and Frommer 1987; Pust et al. 1999; Baylin and Herman 2000; Robertson and Wolffe 2000; Warnecke and Bestor 2000; Feinberg 2001; Jones and Baylin 2002; Cui et al. 2003). There is a growing interest in understanding the correlation between aberrant DNA methylation and tumorigenesis (Huang et al. 1999; Toyota et al. 1999; Costello et al. 2000; Yamashita et al. 2003) In order to facilitate disease marker discovery, diagnostic tool development, and the study of chemotherapeutic response (Laird 2003)

Current methods for detecting 5-methylcytosine can be divided into three major approaches (Laird 2003): (1) profiling methylation globally, (2) identifying methylation patterns at a cluster of CpG sites, and (3) determining methylation levels at individual CpG dinucleotides. Each category offers a different perspective for studying DNA methylation. In general, global screening methods rely on methylation-sensitive restriction enzyme digestion and provide opportunities for new epigenetic marker discovery (Huang et al. 1999; Toyota et al 1999; Costello et al. 2000; Yamashita et al. 2003). Methylation-specific PCR (MSP) (Herman et al 1996) and variations of this procedure (Cottrell and Laird 2003; Zeschnigk et al 2004) were introduced to study the methylation pattern of a few closely neighboring CpG

sites. Since DNA methylation is believed to be an early event during carcinogenesis (Laird 1997), the high sensitivity of MSP is suitable for use as an early detection tool on known epigenetic markers (Huque et al. 2004). For quantitative assessment of individual CpG dinucleotide methylation status, the commonly used methods, including bisulfite treatment, were followed by sequencing (e.g., bisulfite sequencing and pyrosequencing) (Frommer et al. 1992; Uhlmann et al 2002; Dupont et al. 2004; Yang et al 2004), primer extension (e.g., SNuPE) (Gonzalgo and Jones 1997), restriction enzyme digestion (e.g., COBRA) (Xiong and Laird 1997), or real-time PCR (Zeschnigk et al. 2004) These assays provide quantitative profiling or detailed analysis of 5-methylcytosine distribution. The quantitative information generated is currently being used for correlating disease-specific methylation markers to clinical outcomes and facilitating the discovery of anti-tumorigenic drugs (Cheng et al 2004; Issa 2004) However, the current methods analyze CpG methylation status one gene at a time and have limited multiplexing capability. Bisulfite sequencing provides the most comprehensive data on methylation status at every CpG but requires subcloning and sequence analysis of 10-20 individual clones. Higher throughput has been achieved by combining bisulfite-PCR with microarray technology utilizing oligonucleotide probes designed to form a perfect match with either methylated or unmethylated alleles within the target sequences (Adorján et al. 2002; Balog et al. 2002; Gitan et al 2002). This allows parallel evaluation of CpG methylation status at numerous CpG sites across multiple genomic regions of interest. However, bisulfite treatment renders genomic DNA into AT-rich sequences, which exacerbates nonspecific and mismatch hybridizations due to differences in annealing temperatures between different probe sequences. In ad-

[5]Corresponding author.
E-mail barany@med.cornell.edu; fax (212) 746-8104.
Article published online ahead of print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.4181406

dition, probes containing two or more CpG dinucleotides may lack the sensitivity to distinguish partially methylated sequences from those that are fully methylated in heterogeneous clinical samples.

We seek to develop a robust assay for clinical application that provides quantitative methylation levels for multiple CpG dinucleotides in a given genomic region, as well as allowing specific evaluation of many genes in parallel. Such an assay can provide a representational CpG methylation profile of candidate genomic regions, and this profile information may be useful for disease stratification or as predictors of therapeutic response This work presents a new method that aims to substantially improve

quantitative microarray-based methylation detection to meet these needs. As illustrated in Figure 1, combining PCR, ligase detection reaction (LDR), and universal Array (where zip-code sequences appended to LDR primers, guide products to zip-code complements on an array) (Gerry et al 1999; Favis et al. 2000) allows multiplexing and provides high specificity and accuracy. A detailed, quantitative methylation profile of essentially any set of CpG dinucleotides can be determined by using this assay. Fifteen tumor suppressor genes commonly linked to transcriptional silencing in various human cancers were chosen and the methylation status of their promoter regions evaluated (http://www.mdanderson.org/departments/methylation). Up to six CpG dinucleotides per promoter regions were investigated and a total of 75 CpG dinucleotides queried per sample.

Bisulfite/PCR-PCR/LDR/Universal Array
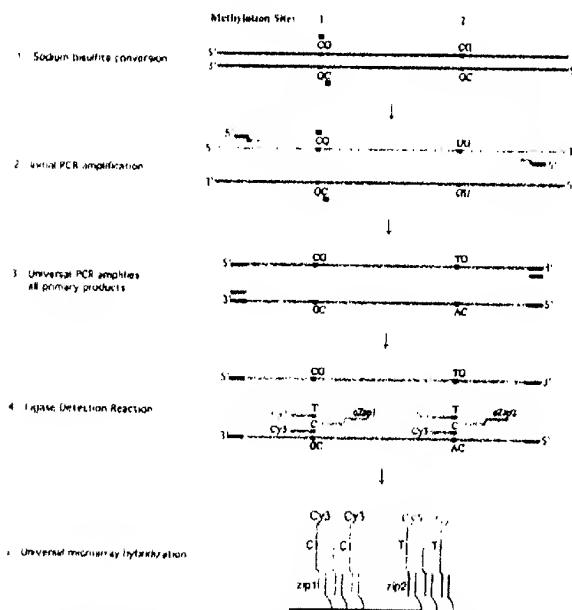
Figure 1. Schematic diagram of the assay. Two hypothetical CpG dinucleotide sites 1 and 2 are designated as methylated and unmethylated, respectively. Sodium bisulfite converts unmethylated, but not methylated, cytosines into uracils. This conversion renders the genomic DNAs into two asymmetrical, noncomplementary strands, and only one designated bisulfite-modified strand (highlighted in orange) is amplified and analyzed. In the initial amplification, PCR primers are designed with a gene-specific 3' portion and an upstream universal sequence (highlighted in black). This universal sequence is used as a PCR primer in the subsequent PCR to simultaneously amplify all the primary amplicons (for ease of illustration, only one amplicon is shown) LDR is performed in a multiplex fashion with three primers (two discriminating and one common primers) interrogating each of the selected CpG sites. The discriminating primers contain a 5' fluorescent label and a 3' discriminating nucleotide to determine either methylated (with 5' Cy3 and 3' C) or unmethylated (with 5' Cy5 and 3' T) cytosines. The common primers bear a 5' phosphate and a 3' unique zip-code complement sequence (e.g., cZip1 and cZip2). Ligation occurs only if the nucleotides at the ligation junction are perfectly base-paired with a complementary template and the ligation products are captured onto a Universal microarray with prespotted zip-codes (addresses). For example, address Zip1 identifies methylated cytosine in methylation site 1, and address Zip2 identifies unmethylated cytosine in methylation site 2. Three Universal microarray addresses are assigned for each promoter region, and each address is double-spotted to ensure the quality of array fabrication and oligonucleotide hybridization efficiency.

Results

Determining assay specificity and quantitative accuracy of bisulfite-PCR / LDR / Universal Array

The general design of the assay is illustrated in Figure 1. Genomic DNAs were treated with sodium bisulfite to convert unmethylated, but not methylated, cytosines into uracils. We have modified the standard bisulfite protocol to ensure a thorough deamination of unmethylated cytosines and increase DNA recovery (Boyd and Zon 2004). Gene-specific PCR primers bearing 5' universal tails were designed to flank each promoter region. A second, universal PCR step allows approximately equal fragment amplification of all sequences amplified in the primary PCR. Since PCR is not the final readout in this assay, primer design is flexible and less constrained by sequence context and is independent of CpG dinucleotide methylation status. Three LDR primers were used to determine the methylation status of each CpG dinucleotide. LDR primers were designed to tolerate mismatched base pairs at internal CpG sites and allow hybridization to fully and partially methylated sequences, as well as unmethylated sequences. A high fidelity Tth ligase (Luo et al. 1996), only ligates the upstream (discriminating) and downstream (common) primers when the 3' discriminating nucleotide at the junction is complementary to the DNA template. This feature allows accurate, quantitative detection of targeted CpG dinucleotides regardless of the presence of internal CpG dinucleotides within the primer sequences. For example (Fig 1), at the methylated CpG site 1, only Cy3-C-labeled ligation products are formed, whereas only Cy5-T-labeled ligation

products are formed at the unmethylated CpG site 2. Unique complementary zip-code sequences on the 3' ends of common primers guide LDR products to their corresponding zip-codes on a Universal Array (Gerry et al 1999; Favis et al 2000). The zip-codes are unique sequences designed with a constant T_m and have no homology to either the target sequence or to other sequences in the genome This design eliminates false signals due to nonspecific binding and mismatch hybridizations

The assay was validated on genomic DNAs extracted from two commonly used colorectal (HCT15) and prostate cancer (LNCaP) cell lines. Promoter regions chosen in this study have included 15 tumor suppressors (CDKN2B [formerly known as p15INK4b], CDKN2A [formerly known as p16INK4a], CDKN2D [formerly known as p14ARF], CDKN1A [formerly known as p21cip], CDKN1B [formerly known as p27kip], TP53 [formerly known as p53], BRCA1, TIMP3, APC, RASSF1, CDH1 [formerly known as ECAD], MGMT, DAPK1 [formerly known as DAPK], GSTP1, and RARB [formerly known as RARβ]) and one hemi-methylated imprinted gene (SNRPN, as an internal control). The methylation profiles of the candidate promoter regions were determined by bisulfite sequencing, which revealed CDKN2B, CDKN2A, CDKN2D, CDKN1A, CDKN1B, TP53, BRCA1, DAPK1, CDH1, MGMT, and TIMP3 were unmethylated in LNCaP, while CDKN2A, CDKN2D, MGMT, RARB, and RASSF1 were methylated in HCT15 among 15 tumor suppressor genes. The initial LDR/Universal Array assay was designed to evaluate methylation status of three CpG sites per promoter region. LDR primers detecting methylated and unmethylated cytosines were validated by using in vitro methylated (SssI methylase) and untreated normal human lymphocyte genomic DNAs, respectively (data not shown) Following bisulfite treatment genomic DNA of each cell line sample was multiplex PCR amplified, and the pooled PCR products were subjected to LDR/Universal Array analysis (Fig 2A) We tested the assay specificity by using LNCaP DNA and subsets of LDR primers that detect only unmethylated cytosines (Fig 2B data not shown) The capture of Cy5 fluorescence signals only at the designated zip-code addresses for each LDR primer set indicated that LDR/Universal Array did not generate nonspecific ligation products and that mismatch hybridization was absent To further demonstrate the assay's accuracy, LDR primers that detected only methylated cytosines were used to investigate a total of 48 CpG sites simultaneously for each cell line (Fig 2C). Our data are consistent with the bisulfite sequencing results, indicating an accurate methylation profile was obtained Different levels of fluorescence intensity were observed at several zip-code addresses. These variations suggested that the targeted CpG dinucleotides may have different methylation levels within the same promoter regions (e g , RASSF1 and TIMP3).

To determine if the assay could be quantitative, genomic DNA from HCT15 (carrying methylated CpG dinucleotides) was mixed with normal human lymphocytes (carrying unmethylated alleles), such that the test samples contained 0%, 20%, 40%, 60%, 80%, and 100% of HCT15 DNA These mixtures were subjected to bisulfite-PCR/LDR/Universal Array analysis (Fig 3A; data not shown) The average fluorescence intensity representing either methylated (Cy3) or unmethylated (Cy5) alleles from each double-spotted zip-code address was used to calculate the methylation ratio of $Cy3/(Cy3 + Cy5)$. Each experiment was repeated at least twice and produced consistent results. Most of the CpG dinucleotides we evaluated have R^2 values between 0.98 and 0.89. Those CpG sites that gave lower R^2 values are likely due to inefficient competition between LDR primers targeted toward
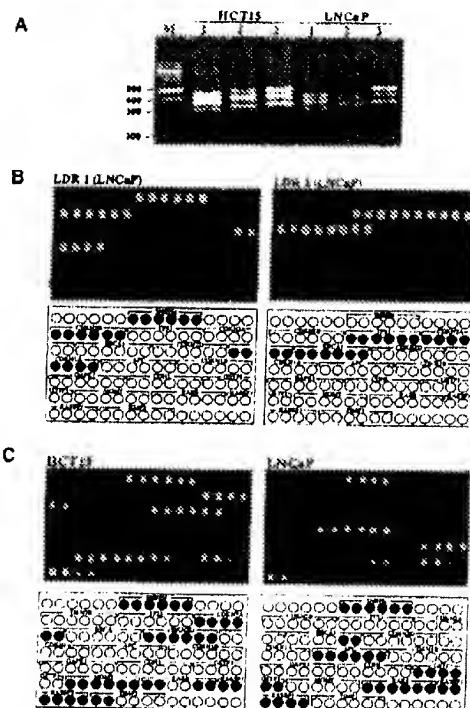
Figure 2. Representative bisulfite-PCR/LDR/Universal Array analysis of 16 promoter regions of cell lines HCT15 and LNCaP. (A) For the ease of demonstration, either five or six promoter regions were amplified in one PCR, and a total of 16 genes were simultaneously analyzed The gene names and the corresponded PCR fragments are as follows: (lane 1) CDKN2B (317 bp), CDKN2A (363 bp), CDKN1A (391 bp), CDKN1B (426 bp), SNRPN (442 bp), and BRCA1 (459 bp); (lane 2) CDKN2D (346 bp), TIMP3 (404 bp), APC (433 bp), RASSF1 (474 bp), and CDH1 (513 bp); and (lane 3) MGMT (362 bp), TP53 (418 bp), DAPK1 (434 bp), GSTP1 (507 bp), and RARB (522 bp). (B) LDR/Universal Array analysis of the unmethylated cytosines in LNCaP amplicons. All PCR products were pooled as LDR templates, but only selected LDR primers were used in each reaction (LDR set1: SNRPN, CDKN2B, CDKN1A; LDR set2: CDKN2A, TP53, BRCA1) The subset of promoter regions that were interrogated in each LDR are depicted in the diagram (green circles) under each array image The Cy5-labeled LDR products (false color green, designed for unmethylated cytosines) were captured on Universal Arrays. (C) All PCR products of each sample were pooled and subjected to LDR/Universal Array assay Only Cy3-labeled LDR primers (false color red) were used in this assay to detect methylated cytosines The diagram under each array image depicts the correlated zip-codes (circles) that were assigned to represent the CpG methylation status in each of the 16 promoter regions Each zip-code was double-spotted on the array to ensure fabrication quality Red and empty circles represent methylated and unmethylated CpG sites, respectively Pink circles represent those CpG dinucleotides that have lower level of methylation The PCR and LDR primer sequences and their concentrations used in these experiments were listed in the Supplemental Tables 1, 2, and 3.

A
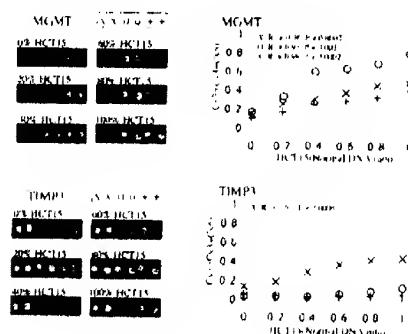
B
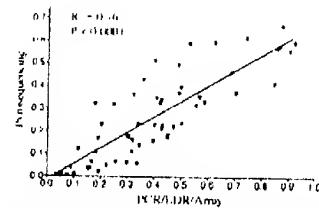
Figure 3. The quantification curves of the assay. Genomic DNAs of HCT15 and normal human lymphocytes were mixed in 0%, 20%, 40%, 60%, 80%, and 100% ratios and subjected to bisulfite-PCR/LDR/ Universal Array analysis. (A) Representative array images are shown scanned in both Cy3 and Cy5 channels. False color red (Cy3) and green (Cy5) represent the methylated and unmethylated alleles of CpG di- nucleotides, respectively. Color composites of the two channels reflect the methylation levels. Each zip-code was double-spotted on the array to ensure fabrication quality. MGMT and TIMP3 were used as examples to show the assay linearity measured at individual CpG dinucleotides. The plotted value at y axis represents the fluorescence intensity Cy3/ (Cy3 + Cy5) ratio. The value at x axis represents the percentage of HCT15 mixed with normal human lymphocyte genomic DNAs. The R² and P. values of each linear regression line were calculated, although the lines were omitted in the plots for visual clarity. Nearly no methylation was observed at two of the CpG sites of TIMP3 resulting in poor statistical correlation (circles R² = 0.81, P = 0.02, crosses R² = 0.08, P = 0.01). The experiments were repeated three times with different sample prepara- tions and array hybridizations. The PCR and LDR primer sequences and their concentrations used in these experiments are listed in Supplemental Tables 3 and 4. (B) Comparison of the MGMT methylation level from 15 colorectal carcinomas using pyrosequencing technology and bisulfite/ PCR/LDR/Universal Array. Three CpG sites were evaluated per DNA sample. The plotted value at y-axis represents the percentage of meth- ylated cytosines in the tumor samples as obtained from pyrosequencing. The value at x-axis represents the ratio of fluorescence intensities Cy3/ (Cy3 + Cy5). The mixed genomic DNAs of HCT15 and normal human lymphocytes shown in Figure 3A were included as controls.

unmethylated and methylated alleles and can be resolved by re- designing the LDR primers to have a higher melting temperature. Our analysis confirmed the different percentage of methylation at each CpG dinucleotide and suggested that methylation level is not 100% at each CpG site in tumor cell line DNA. For example,

for HCT15, a medium level of methylation was observed at the first CpG dinucleotide and <10% methylation at the other two CpG dinucleotides in TIMP3. By comparing the ratio of (meth- ylated)/(methylated + unmethylated) DNA in different cell lines, we could extrapolate the CpG methylation level at a given posi- tion. Alternatively, SssI could be used to methylate DNA in vitro to completion to generate standard curves for calibration (data not shown). We tested assay sensitivity by mixing tumor cell line DNA with normal human lymphocyte DNAs. A preliminary study suggested that the unbiased PCR primer design was suffi- cient to detect the presence of methylated alleles, even when diluted down to 1% in unmethylated alleles (Supplemental Fig 1). Nevertheless, in the colorectal cancer study, we currently use 10%–15% as a cut-off for our scoring criteria. Overall, our data demonstrate that bisulfite-PCR/LDR/Universal Array approach is a quantitative and sensitive method for the measurement of DNA methylation.

Comparison of the method with a pyrosequencing assay using clinical tumor samples

To further evaluate the quantitative accuracy and clinical utility, we analyzed MGMT methylation level at three CpG dinucleotides per sample on a total of 15 colorectal tumors, using both our assay as well as by pyrosequencing. Genomic DNAs of these tu- mors were bisulfite treated, amplified by using multiplex PCR, and analyzed by LDR/Universal Array. Alternatively, MGMT was uniplex amplified from the same tumor DNAs for pyrosequenc- ing. Three sequencing primers were designed to investigate the methylation level of the same CpG dinucleotides that were stud- ied by LDR/Universal Array. The comparison revealed a high cor- relation between these two methods (Fig 3B). The few samples where results varied may be due to variations in array fabrication, differences in efficiency of multiplex versus uniplex PCR ampli- fication, and, most likely, the pyrosequencing primer design. Two of the sequencing primers contained two and three CpG sites, respectively, and generated a biphasic plot while establish- ing standard curves. This may have led to bias in the quantifica- tion seen by pyrosequencing. Nevertheless, this highly signifi- cant correlation suggests bisulfite/PCR/LDR/Universal Array is an accurate method for quantitative analysis of heterogeneous clini- cal samples.

Application of the method to cancer cell lines and heterogeneous clinical samples

We performed a blinded study to determine methylation profiles of several cancer cell lines (five colorectal, one breast, and one prostate) with this approach (Fig. 4). Three CpG sites per pro- moter region were evaluated. Methylation was not observed in the promoter regions of CDKN2B, CDKN1A, CDKN1B, TP53, and BRCA1 among all the tested cell line DNAs. These results sug- gested that promoter methylation is not a random event. All cell lines under the blinded study have very distinct methylation profiles among the 15 tumor suppression genes except for two pairs of cell lines: HCT15/DLD-1 and HT29/WiDr, which share almost identical promoter methylation patterns. Each pair of cell lines HCT15/DLD-1 and HT29/WiDr was established from the same colon carcinomas (additional cell line information is de- scribed in American Type Culture Collection [ATCC] Web site http://www.atcc.org) (Chen et al. 1995). This observation sug- gested that colorectal cancer cell lines derived from the same tumor have essentially identical methylation profiles that are
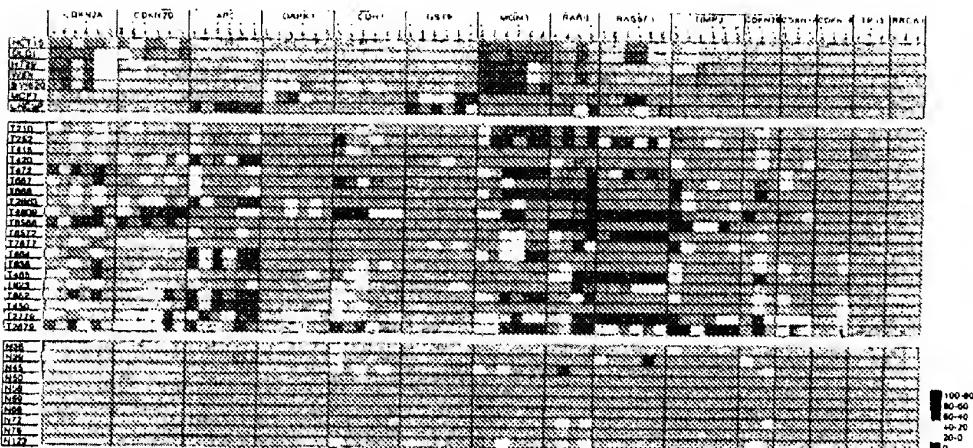
Figure 4. Selected examples of DNA methylation profiles of cancer cell lines and clinical samples. Five colorectal (HCT15, DLD-1, HT29, WiDr, and SW620), one breast (MCF7), one prostate (LNCaP) cancer cell lines, 20 primary colorectal cancer (T series), and 10 adjacent normal tissues (N series) were analyzed. Six CpG sites per promoter region were analyzed in each sample except CDKN2B, CDKN1A, CDKN1B, TP53, and BRCA1. Standard curves were not established for the genes mentioned above since hypermethylation is not reported in the literature or observed in our clinical sample study. The standard curves should be established when applying this assay to other tumor types such as breast cancer. Around 10%–15% established CpG methylation standard curves gave lower R² values (between low 80s and 70s), including CDKN2A (CpG_3 and CpG_6), CDKN2D (CpG_1), GSTP1 (CpG_3), DAPK1 (CpG_3), RASSF1 (CpG_4), and TIMP3 (CpG_2). Only four CpG sites of RARB were shown due to defects on zip-code addresses during array fabrication. The color scale represents the percentage of methylation levels determined from the standard curves. Notice that for the cell lines HCT15/DLD-1 and HT29/WiDr, each pair is derived from a same tissue origin reflected in their identical methylation patterns. The PCR and LDR primer sequences and their concentrations used in these experiments are listed in the Supplemental Tables 3, 4, and 5.

distinct from cell lines derived from a different patient. To ensure the accuracy of scoring a candidate gene promoter as hypermethylated, the methylation status of three additional CpG sites per promoter was examined (examples are shown in Supplemental Fig. 2) for the 10 genes that were found to be methylated in the cell line study (CDKN2A, CDKN2D, TIMP3, APC, RASSF1, CDH1, MGMT, DAPK1, GSTP1, and RARB). Within each promoter, the six CpG sites examined were dispersed evenly throughout the amplified genomic sequence. Some of the CpG sites investigated were within 50 bases and resulted in the design of overlapping LDR primers. Nevertheless, we have found that LDR efficiency was not altered by the overlapping primer design (data not shown). Thus, the cytosines of all 75 CpG sites for a given sample were interrogated independently at the same time.

In a pilot study, we have profiled methylation status of 96 colorectal tumor samples and 73 matched normal tissues (selected data are shown in Fig 4; Supplemental Fig. 2). Our preliminary analysis shows a rich variety of methylation profiles. Seven promoter regions (CDKN2A, CDKN2D, APC, MGMT, RARB, RASSF1, and TIMP3) showed statistically significant increased methylation in tumor compared to normal tissues. Promoter regions of CDKN1A, CDKN1B, TP53, and BRCA1 revealed little or no methylation. Thus, methylation profiles of clinical samples were similar to those observed in the colorectal cancer cell lines. One of the CpG dinucleotides of CDKN2B was frequently methylated in the tumor samples. The biological significance of this methylation remains to be investigated by examining additional CpG dinucleotides to determine CDKN2B promoter hypermethylation status. Nevertheless, these results reaffirm that methyl-

ation in colorectal cancer is gene specific and nonrandom. Additional tumor samples will be examined to provide sufficient power to determine the correlation between promoter methylation and the tissue pathological-clinical information.

Discussion

We have presented an accurate and quantitative assay that provides a representational CpG methylation profile from colorectal cancer samples, where stromal cell infiltration is often seen. This assay simultaneously determines the DNA methylation status of multiple gene promoters, querying a total of 75 CpG dinucleotide sites per sample. Genomic DNAs isolated from seven cancer cell lines and 169 colon tumor samples were tested. Cell lines derived from the same tumor have essentially identical methylation profiles. The percentage of CpG dinucleotide methylation of MGMT promoter in clinical samples was compared by our assay with those derived from pyrosequencing and resulted in a high correlation.

The candidate promoter regions, involving genes in cell cycle regulation, DNA repair, and tumor metastasis and invasion, were previously reported in the literatures as associated with abnormal gene silencing in tumors or cancer cell lines. Although the percentage of methylated promoters in a cohort may vary due to the sample source and the assays used in determining DNA methylation status, our preliminary analysis of the colorectal tumor methylation profile gave consistent results as those published previously (Esteller et al. 2001). For example, in agreement with our own findings, it has been shown that there is

essentially no hypermethylation in *CDKN2B*, *TP53*, *BRCA1*, and *GSTP1* promoter regions, while *CDKN2D*, *CDKN2A*, *MGMT*, and *APC* were methylated to a higher level among all the samples tested (Esteller et al., 2001).

Our assay allows virtually any CpG site in the promoter and first intron regions for more than a dozen genes to be analyzed simultaneously. There is increasing evidence that genes such as *MLH1* and *RASSF1A* exhibit an increasing gradient of methylation from the promoter proximal region to the first exon (Deng et al. 1999; Yan et al. 2003). To avoid the bias of scoring a hyper (or hypo)-methylated promoter and linking it to its disease state, multiple CpG sites across a larger window of the genomic region should be investigated in each assay. Studies done with MSP- and restriction enzyme-based methods only reveal the methylation pattern of small sequence regions; additional sequence contexts may be needed to sufficiently determine the promoter methylation status. Our approach interrogated multiple CpG dinucleotides that were evenly distributed over 300–500 bases. For paraffin-embedded tissue, two or more adjacent shorter PCR amplicons should be designed to overcome the poor amplification typically observed in these types of samples (Supplemental Fig. 3). Moreover, the LDR efficiency was not affected, even when some of the LDR primers were designed with overlapping sequences. This technique provides a detailed mapping of the methylation profile in each promoter CpG island locus that may correlate with transcriptional silencing during disease progression.

The bisulfite-PCR/LDR/Universal Array approach provides several advantages over existing methods for the analysis of DNA methylation patterns. First, there are two levels of specificity facilitated by gene-specific primers, initially during PCR and subsequently during LDR. Given the numerous duplications annotated and suspected in the human genome, this approach enhances the ability to reliably target the CpG islands in the locus of interest. Second, LDR allows accurate identification of low abundance nucleotide alterations with a remarkable accuracy due to the high fidelity of Tth ligase (Luo et al. 1996). This unique feature eliminates the concern of mismatch hybridization due to partial methylation of internal CpGs of the LDR primer sequences and allows LDR primers to be placed at essentially any CpG dinucleotide of interest. Third, the unique zip-code sequences are designed with similar T_m across the platform and have no sequence homology in the human genome. This feature allows only ligated LDR products to be captured, thus avoiding background signals. As the Universal Array can be easily expanded (Gerry et al 1999), additional CpG sites or genes can be evaluated in a single assay. Fourth, this assay has the potential to detect low abundance methylated alleles. By redesigning the multiplex gene-specific PCR primers to be methyl specific, a detection of at least 0.1% was achieved, albeit with reduced quantitative dynamic range (Supplemental Fig. 1). An inherent problem with many DNA amplification techniques is that greater detection sensitivity comes at the cost of increased false positives. Although methylation assays relying solely on PCR as a readout tool offer sensitive detection, they are prone to false positives resulting from the AT richness or incomplete deamination of bisulfite-treated DNAs. Consequently, such assays would be limited in their multiplex capability. Our assay confirmed promoter methylation status via six CpG sites within a PCR fragment; this approach offers high specificity and accuracy while avoiding false positives. Moreover, each module of the PCR/LDR/Universal microarray approach is ideal for multiplexing and can be automated by using a liquid handling system to increase throughput.

A recent publication has reported the detection of dozens to hundreds of possible mutations by using multiplex PCR/LDR in a single-tube format (Favis et al. 2004). The capture of multiplex LDR products onto an array format provides an efficient "modular" readout and substantially increases assay throughput. Universal microarray experiments in our laboratory are now performed in an array-of-arrays format, where 64 array hybridizations are carried out simultaneously (data not shown). This array-of-arrays approach drastically reduces the cost of array fabrication, minimizes the variation during hybridization, and increases throughput.

In summary, we present a robust and accurate method that determines cytosine methylation at any selected set of CpG dinucleotides in the genome. Importantly, this new method allows the evaluation of methylation level at individual CpG sites. Quantitative values for this parameter may facilitate stratification of tumors, since based on the degree of methylation, it may be possible to estimate disease progression. In addition, the ability to quantify methylation at specific sites in advanced tumors may provide information on tumor heterogeneity. These data will enable clinical decisions related to individualized treatment strategies. Our goal is to expand this prototype assay into a focused array platform that investigates 30–50 frequently methylated tumor suppressor promoters in several different human carcinomas. Around 10–15 individual CpG dinucleotides evenly distributed over a larger window of each promoter region will be interrogated. We anticipate that such a focused platform will facilitate the development of DNA-based molecular markers for disease diagnosis and prognosis and will be suitable for routine clinical use.

Methods

Cell line culture, tumor samples, and DNA extraction

Normal human lymphocyte genomic DNA was purchased from Roche. Colorectal, breast, and prostate cancer cell lines were obtained from American Type Culture Collection and cultured under the ATCC-recommended media conditions. Fresh frozen primary colorectal adenocarcinomas were obtained from Memorial Sloan Kettering Cancer Center under Institutional Review Board (IRB)-approved protocols. Genomic DNAs were extracted by using the DNeasy Tissue Kit (Qiagen) according to the manufacturer's guidelines.

Sodium bisulfite treatment of genomic DNAs

Typically, 1 µg genomic DNA was denatured in 40 µL of 0.2 N NaOH by incubating for 10 min at 37°C before addition of 30 µl of freshly prepared 10 mM hydroquinone and 520 µL of 3 M sodium bisulfite. The reaction was incubated for 20 min at 50°C, 15 sec at 85°C for 48 cycles (16 h). The DNA clean-up procedure was as follows: (1) the total reaction volume (~600 µL) was transferred to a Microcon NCO30 filter (Millipore) and centrifuged at 13000g for 16 min; (2) 500 µL deionized H_2O were added to the upper chamber, centrifuged at 13,000g for 7 min, the filtrate discarded, and the wash repeated twice; (3) 500 µL of 0.3M NaOH were added to the upper chamber, incubated for 5 min at room temperature and then centrifuged at 13,000g for 8 min; (4) 500 µl deionized H_2O were added to the upper chamber, centrifuged at 13,000g for 8 min, the filtrate discarded, and the wash repeated; and (5) the filter was inverted to collect the bisulfite-converted DNA. An appropriate volume of water (if needed) was

used to rinse the upper chamber to recover DNA in a final volume of 20 µl

Multiplex PCR amplification

The multiplex PCR consists of two stages. PCR stage I (12.5 µL) contained 1.5 µL bisulfite-modified DNA, 400 µM of each dNTP, 1× AmpliTaq Gold PCR buffer, 4 mM MgCl₂, and 1.25 U AmpliTaq Gold polymerase (Applied Biosystems). Mineral oil was added prior to thermal cycling. The PCR stage I conditions were as follows: 10 min at 95°C; 15 cycles of 30 sec at 94°C, 1 min at 60°C, and 1 min at 72 °C; followed by a final extension step of 5 min at 72°C. PCR stage II (12.5 µL) contained 400 µM of each dNTP, 1× AmpliTaq Gold PCR buffer, 4 mM MgCl₂, 12.5 pmol universal primer (UniB2, see Supplemental Table 1), and 1.25 U AmpliTaq Gold polymerase. The 12.5 µL reaction mixture was added through the mineral oil to the completed stage I PCR. The PCR stage II conditions were as follows: 10 min at 95°C; 30 cycles of 30 sec at 94°C, 1 min at 55°C, 1 min at 72°C; followed by a final extension step of 5 min at 72°C. Taq DNA polymerase was inactivated by adding 1.25 µl Proteinase K (20 mg/mL, Qiagen) to the completed stage II PCR, incubating for 10 min at 70°C and 15 min at 90°C. Before pooling the PCR products for LDR assay, the presence of amplicons was confirmed by electrophoresis on a 3% agarose gel.

LDR, Universal Array hybridization, and data analyses

A typical LDR (20 µL) contained 20 mM Tris-HCl (pH 7.6), 10 mM MgCl₂, 100 mM KCl, 10 mM DTT, 1 mM NAD, 25 fmol wild-type Tth ligase (Zirvi et al 1999), 500 fmol of each LDR primer, and 5–10 ng of each PCR amplicon. The LDR conditions were as follows: 3 min at 95°C; 25 cycles of 30 sec at 95°C and 4 min at 60°C. The LDR reaction was diluted with an equal volume of 2× hybridization buffer (600 mM MES at pH 6.0, 20 mM MgCl₂, 0.2% SDS), denatured for 3 min at 95°C and plunged on ice. The Universal Arrays were pre-equilibrated with 1× hybridization buffer at room temperature for at least 15 min. Coverwells (Grace Bio-labs) were attached to arrays and filled with 40 µl denatured LDR reactions. The assembled arrays were incubated in a rotating hybridization oven for 60 min at 65°C. After hybridization, the arrays were washed in 300 mM Bicine (pH 8.0), 0.1% SDS (or 10 min at 60°C. An updated version with 384 addresses will accommodate all the LDR products. Each array was scanned by using a Perkin Elmer ProScanArray under the same laser power and PMT within the linear dynamic range. The Cy3 and Cy5 dye bias was determined by measuring the fluorescence intensity of an equal mole of Cy3- and Cy5-labeled LDR primers manually deposited on a slide surface. This fluorescence intensity ratio (W = I_{Cy3}/I_{Cy5}) was used to normalize the label bias when calculating the methylation ratio Cy3/(Cy3 + Cy5). MetaMorph Imaging System (Universal Imaging) was used to create images depicting the Cy3 (red) and the Cy5 (green).

Oligonucleotide design and synthesis

Oligonucleotides were obtained from IDT or synthesized in-house on an ABI 394 DNA Synthesizer (PE Biosystems) using standard phosphoramidite chemistry (Khanna et al. 1999). Spacer phosphoramidite C18, 3'-amino-modifier C3 CPG, C3 spacer, and Cy3, Cy5, and standard phosphoramidites were purchased from Glen Research. All other reagents were purchased from PE Biosystems. The zip-code oligonucleotides were synthesized on a 3' amino modifier C3 column with a spacer C18 inserted before the first base. The common LDR primers were synthesized with 5' phosphates and 3' C3 spacers as blocking groups. Oligonucleotides with cyanine labels were cleaved from

the CPG supports and deprotected according to manufacturer's recommendations. Both labeled and unlabeled LDR oligonucleotides were purified and desalted on SuperPore columns (Biosearch Technologies) according to the manufacturer's instructions, then spin-dried (Speed-Vac) and stored at −20°C. For those primers that inevitably covered CpG dinucleotides in the body of their sequences, the nucleotides that base paired with cytosines in CpG dinucleotide were synthesized in two ways. One was to use nucleotide analogs dK or dP in the primers' syntheses. The pyrimidine derivative dP base pairs with either A or G, while the purine derivative dK base pairs with either C or T at similar efficiency. Alternatively, to reduce the cost of primer synthesis, those nucleotide positions with analogs dK or dP incorporated were substituted by nucleotides dG or dC, respectively. For example, the substituted nucleotide dG in a PCR primer formed either Watson-Crick base pair with C (methylated) or wobble base pair with U (unmethylated) on the bisulfite-modified DNA template.

Universal Array fabrication

Polymer-coated slides were fabricated as previously described (Gerry et al 1999, Favis et al 2000) or were purchased (CodeLink slides) from Amersham Biosciences. Universal Arrays were spotted by using a Pixsys5500 robot with a quill-type spotter in a controlled humidity chamber (Cartesian Technologies). Zip-code oligonucleotides each with a unique 24-mer sequence were prepared by mixing 5 µl of 1000 µM stock oligonucleotides with 5 µl of 0.4 M K₂HPO₄/KH₂PO₄ (pH 8.5) in 384 conical well spotting plates. Arrays were printed under relative humidity 60%–70%. To ensure that all the zip-codes were spotted without cross-contamination during array fabrication, one out of 10 slides on average was subjected to quality control by hybridizing fluorescein-labeled zip-code complements targeting a combination of rows or columns of zip-code addresses. A batch of fabricated arrays passed the quality control only when specific fluorescence signals were present on all the targeted rows and columns without extraneous signals on the adjacent, unexpected neighboring addresses.

Pyrosequencing

A promoter sequence of MGMT was PCR amplified by using 1 µL bisulfite-modified DNA, 400 µM of each dNTP, 1× AmpliTaq Gold PCR buffer, 4 mM MgCl₂, 0.2 µM PCR primers (5'-GGTTTTAGGAGGGGAGAGATT-3' and 5'-CCTAACCCRA ATAACCCTTC-3'), and 1.25 U AmpliTaq Gold polymerase (Applied Biosystems). The PCR condition was as follows: 15 min at 94°C; 45 cycles of 15 sec at 95°C, 30 sec at 58°C, and 15 sec at 72°C, followed by a final extension step for 5 min at 72°C. Three sequencing primers (5'-GTAGTAGTTTTAGAGTAGGAT-3', 5'-TTTAGAGAGTTTTTAGGAT-3' and 5'-AAATTAAGGTATA GAGTTTT-3') were designed to determine the CpG dinucleotide methylation levels. The primers were designed and experiments were performed by Biotage.

References

Adorján, P., Distler, J., Lipscher, E., Model, F., Meuller, J., Pelet, C., Braun, A., Florl, A.R., Gütig, D., Grabs, G., et al. 2002. Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.* **30:** e21.

Balog, R.P., de Souza, Y.E., Tang, H.M., DeMasellis, G.M., Gao, B., Avila, A., Gaban, D.J., Mittelman, D., Minna, J.D., Luehke, K.J., et al. 2002. Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays. *Anal. Biochem.* **309:** 301–310.

Baylin, S.B and Herman, J.G. 2000. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* **16:** 168–174.

Boyd, V.L. and Zoo, G. 2004. Bisulfite conversion of genomic DNA for methylation analysis: Protocol simplification with higher recovery applicable to limited samples and increased throughput. *Anal. Biochem.* **326:** 278–280.

Chen, T.R., Dorotinsky, C.S., McGuire, L.J., Macy, M.L., and Hay, R.J. 1995. DLD-1 and HCT-15 cell lines derived separately from colorectal carcinomas have totally different chromosome changes but the same genetic origin. *Cancer Genet. Cytogenet.* **81:** 103–108.

Cheng, J.C., Yoo, C.B., Weisenberger, D.J., Chuang, J., Wozniak, C., Liang, G., Marquez, V.E., Greer, S., Orntoft, T.F., Thykjaer, T., et al. 2004. Preferential response of cancer cells to zebularine. *Cancer Cell* **6:** 151–158.

Costello, J.F., Fruhwald, M.C., Smiragha, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C., et al. 2000. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.* **24:** 132–138.

Cottrell, S.E. and Laird, P.W. 2003. Sensitive detection of DNA methylation. *Ann. N.Y. Acad. Sci.* **983:** 120–130.

Cui, H., Cruz-Correa, M., Giardiello, F.M., Hutcheon, D.F., Kafonek, D.R., Brandenburg, S., Wu, Y., He, X., Powe, N.R., and Feinberg, A.P. 2003. Loss of IGF2 imprinting: A potential marker of colorectal cancer risk. *Science* **299:** 1753–1755.

Deng, G., Chen, A., Hong, J., Chae, H.S., and Kim, Y.S. 1999. Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. *Cancer Res.* **59:** 2029–2033.

Dupont, J.M., Tost, J., Jammes, H., and Gut, I.G. 2004. De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal. Biochem.* **333:** 119–127.

Esteller, M., Corn, P.G., Baylin, S.B., and Herman, J.G. 2001. A gene hypermethylation profile of human cancer. *Cancer Res.* **61:** 3225–3229.

Favis, R., Day, J.P., Gerry, N.P., Phelan, C., Narod, S., and Barany, F. 2000. Universal DNA array detection of small insertions and deletions in BRCA1 and BRCA2. *Nat. Biotechnol.* **18:** 561–564.

Favis, R., Huang, J., Gerry, N.P., Culliford, A., Paty, P., Soussi, T., and Barany, F. 2004. Harmonized microarray/mutation scanning analysis of TP53 mutations in undissected colorectal tumors. *Hum. Mutat.* **24:** 63–75.

Feinberg, A.P. 2001. Methylation meets genomics. *Nat. Genet.* **27:** 9–10.

Feinberg, A.P. and Vogelstein, B. 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301:** 89–92.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci.* **89:** 1827–1831.

Gardiner-Garden, M. and Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196:** 261–282.

Gerry, N.P., Witowski, N.E., Day, J., Hammer, R.P., Barany, G., and Barany, F. 1999. Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* **292:** 251–262.

Gitan, R.S., Shi, H., Chen, C.M., Yan, P.S., and Huang, T.H. 2002. Methylation-specific oligonucleotide microarray: A new potential for high-throughput methylation analysis. *Genome Res.* **12:** 158–164.

Gonzalgo, M.L. and Jones, P.A. 1997. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* **25:** 2529–2531.

Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D., and Baylin, S.B. 1996. Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci.* **93:** 9821–9826.

Hoque, M.O., Begum, S., Topaloglu, O., Jeronimo, C., Mambo, E., Westra, W.H., Califano, J.A., and Sidransky, D. 2004. Quantitative detection of promoter hypermethylation of multiple genes in the tumor, urine, and serum DNA of patients with renal cancer. *Cancer Res.* **64:** 5511–5517.

Huang, T.H., Perry, M.R., and Laux, D.E. 1999. Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.* **8:** 459–470.

Issa, J.P. 2004. CpG island methylator phenotype in cancer. *Nat. Rev. Cancer* **4:** 988–993.

Jones, P.A. and Baylin, S.B. 2002. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3:** 415–428.

Khanna, M., Park, P., Zirvi, M., Cao, W., Picon, A., Day, J., Paty, P., and Barany, F. 1999. Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumors. *Oncogene* **18:** 27–38.

Laird, P.W. 1997. Oncogenic mechanisms mediated by DNA methylation. *Mol. Med. Today* **3:** 223–229.

——. 2003. The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* **3:** 253–266.

Luo, J., Bergstrom, D.E., and Barany, F. 1996. Improving the fidelity of *Thermus thermophilus* DNA ligase. *Nucleic Acids Res.* **24:** 3071–3078.

Post, W.S., Goldschmidt-Clermont, P.J., Wilhide, C.C., Heldman, A.W., Sussman, M.S., Ouyang, P., Milliken, E.E., and Issa, J.P. 1999. Methylation of the estrogen receptor gene is associated with aging and atherosclerosis in the cardiovascular system. *Cardiovasc. Res.* **43:** 985–991.

Robertson, K.D. and Wolffe, A.P. 2000. DNA methylation in health and disease. *Nat. Rev. Genet.* **1:** 11–19.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. 1999. CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci.* **96:** 8681–8686.

Uhlmann, K., Brinckmann, A., Tohat, M.R., Ritter, H., and Nürnberg, P. 2002. Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* **23:** 4072–4079.

Warnecke, P.M. and Bestor, T.H. 2000. Cytosine methylation and human cancer. *Curr. Opin. Oncol.* **12:** 68–73.

Xiong, Z. and Laird, P.W. 1997. COBRA: A sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* **25:** 2532–2534.

Yamashita, K., Dai, T., Dai, Y., Yamamoto, F., and Perucho, M. 2003. Genetics supersedes epigenetics in colon cancer phenotype. *Cancer Cell* **4:** 121–131.

Yan, P.S., Shi, H., Rahmatpanah, F., Hsiau, T.H., Hsiau, A.H., Leu, Y.W., Liu, J.C., and Huang, T.H. 2003. Differential distribution of DNA methylation within the RASSF1A CpG island in breast cancer. *Cancer Res.* **63:** 6178–6186.

Yang, A.S., Estécio, M.R., Doshi, K., Kondo, Y., Tajara, E.H., and Issa, J.P. 2004. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res.* **32:** e38.

Zeschnigk, M., Böhringer, S., Price, E.A., Onadim, Z., Masshöfer, L., and Lohmann, D.R. 2004. A novel real-time PCR assay for quantitative analysis of methylated alleles (QAMA): Analysis of the retinoblastoma locus. *Nucleic Acids Res.* **32:** e125.

Zirvi, M., Bergstrom, D.E., Saurage, A.S., Hammer, R.P., and Barany, F. 1999. Improved fidelity of thermostable ligases for detection of microsatellite repeat sequences using nucleoside analogs. *Nucleic Acids Res.* **27:** e4.

Exhibit 12:   Das et al., "Detection and Serotyping of Dengue Virus in Serum Samples by Multiplex Reverse Transcriptase PCR-Ligase Detection Reaction Assay," *J. Clin. Microbiol.* 46(10):3276–84 (2008)

Detection and Serotyping of Dengue Virus in Serum Samples by Multiplex Reverse Transcriptase PCR–Ligase Detection Reaction Assay[∇][†]

S. Das,[1] M. R. Pingle,[2] J. Muñoz-Jordán,[3] M. S. Rundell,[2] S. Rondini,[1] K. Granger,[2] G.-J. J. Chang,[4]
E. Kelly,[5] E. G. Spier,[6] D. Larone,[7] E. Spitzer,[8] F. Barany,[2] and L. M. Golightly[1]*

*Department of Medicine, Division of International Medicine and Infectious Diseases,[1] Department of Microbiology and Immunology,[2] and
Department of Pathology and Laboratory Medicine,[7] Weill Medical College of Cornell University, New York, and Department of
Pathology, Stony Brook University Medical Center, Stony Brook,[8] New York; Centers for Disease Control and Prevention, San Juan,
Puerto Rico[3], Centers for Disease Control and Prevention, Fort Collins, Colorado[4], Walter Reed Army Institute of Research,
Silver Spring, Maryland[5]; and Applied Biosystems, Foster City, California[6]*

The detection and successful typing of dengue virus (DENV) from patients with suspected dengue fever is important both for the diagnosis of the disease and for the implementation of epidemiologic control measures. A technique for the multiplex detection and typing of DENV serotypes 1 to 4 (DENV-1 to DENV-4) from clinical samples by PCR-ligase detection reaction (LDR) has been developed. A serotype-specific PCR amplifies the regions of genes C and E simultaneously. The two amplicons are targeted in a multiplex LDR, and the resultant fluorescently labeled ligation products are detected on a universal array. The assay was optimized using 38 DENV strains and was evaluated with 350 archived acute-phase serum samples. The sensitivity of the assay was 98.7%, and its specificity was 98.4%, relative to the results of real-time PCR. The detection threshold was 0.017 PFU for DENV-1, 0.004 PFU for DENV-2, 0.8 PFU for DENV-3, and 0.7 PFU for DENV-4. The assay is specific; it does not cross-react with the other flaviviruses tested (West Nile virus, St. Louis encephalitis virus, Japanese encephalitis virus, Kunjin virus, Murray Valley virus, Powassan virus, and yellow fever virus). All but 1 of 26 genotypic variants of DENV serotypes in a global DENV panel from different geographic regions were successfully identified. The PCR-LDR assay is a rapid, sensitive, specific, and high-throughput technique for the simultaneous detection of all four serotypes of DENV.

The dengue virus (DENV), a mosquito-borne flavivirus, consists of four closely related but genetically distinct antigenic serotypes: DENV serotype 1 (DENV-1), DENV-2, DENV-3, and DENV-4. It is tropical and subtropical in distribution and is prevalent in Asia, Africa, and Central and South America (45). Infection with any of the four serotypes of DENV may cause a mild febrile illness, dengue fever (DF). In some cases, however, more-severe manifestations, such as dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS), occur; these may prove fatal without proper early intervention (15).

Geographic spread of both the mosquito vector and the virus over the past 25 years has led to the increased occurrence of epidemic DF/DHF/DSS, making dengue a major global health problem. The disease is endemic in more than 100 countries, with an estimated 2.5 billion people at risk of infection. It is estimated that 50 million DENV infections occur each year, with 500,000 cases of DHF and at least 22,000 deaths, mainly in children (31, 32, 45; WHO/WPRO/SEARO meeting on DengueNet implementation in Southeast Asia and the

Western Pacific, Kuala Lumpur, Malaysia, 11 to 13 December 2003).

DENV infection confers lifelong serotype-specific immunity. Multiple infections with different DENV serotypes occur in regions of hyperendemicity (31, 35). Secondary infections with a different DENV serotype are major risk factors for DHF and DSS (13, 14, 39) due to antibody-dependent enhancement of disease (35). Serotype identification and the differentiation of primary and secondary infections are therefore important both for patient management and for the implementation of public health measures (26, 33).

The diagnosis of DENV infection and the typing of DENV serotypes can be confirmed using viral isolation techniques, serology, or molecular methods. Virus isolation is the gold standard for detection but requires 7 to 10 days and is often insensitive (26). Serological tests for the detection of viral antibodies, such as immunoglobulin M and immunoglobulin G antibody capture enzyme-linked immunosorbent assays, require the demonstration of a rise in antibody titer from an acute-phase to a convalescent-phase serum sample and therefore have little impact on patient management (24, 41). Additionally, the extensive antigenic cross-reactivity in serological assays, both among flaviviruses and between DENV serotypes, further complicates definitive diagnosis and the interpretation of the assays (18, 20, 41).

Molecular techniques based on the detection of genomic sequences by reverse transcription-PCR (RT-PCR), nested PCR, and real-time PCR are rapid and sensitive and have

* Corresponding author. Mailing address: Division of International Medicine and Infectious Diseases, Weill Medical College of Cornell University, 1300 York Avenue, Room A 421, New York, NY 10021. Phone: (212) 746-6320. Fax: (212) 746-8675. E-mail: lgolight@med.cornell.edu

replaced virus isolation as the new standard method for the detection of DENV in acute-phase serum samples (15). These methods identify the four different serotypes by using genus- or serotype-specific primers or a combination of both. A two-step nested RT-PCR approach is routinely employed in laboratories worldwide (27).

Although most molecular techniques have the advantage of being rapid and sensitive, there are limitations. Real-time assays are limited by the large number of reactions required, as in Sybr green-based assays, and the manipulation of samples necessary for serotype identification is a limitation of fluorogenic dye-based assays (20, 26, 29, 44). In addition, the genetic diversity among DENV isolates raises concerns regarding false-negative PCR results due to mismatches in sequences resulting from the continual evolution of variant viral sequences. For example, recent studies indicate the presence of well-defined phylogenetic groups within each serotype of DENV. The genotypes described within the different serotypes are based on sequence variations in gene E and NS1. The number of genotypes varies, ranging from three (for DENV-4) to five (for DENV-1, -2, and -3), depending on the region sequenced (19, 31).

The present study was conducted on clinical samples from Puerto Rico, which is regarded as a prototype of the urban establishment of DENV. During the past 2 decades, Puerto Rico has experienced increasingly severe DENV epidemics (4). All four serotypes and multiple genotypes have been in circulation on this island. This situation is considered ideal for the genetic evolution of the virus, which has been demonstrated by molecular analysis of DENV-2 and -4 (3, 4). The situation is expected to be further complicated by the recent introduction of West Nile virus, which may clinically mimic DF and is difficult to distinguish from DENV by serologic tests, due to cross-reactive antibodies (7, 25).

In this study, we combine multiplex PCR using multiple degenerate primers with a ligase detection reaction (LDR) and a universal zip-code array for the simultaneous identification and serotyping of DENV from viral cultures and clinical samples from Puerto Rico. Originally developed for discriminating single-base mutations or polymorphisms in cancer genes, LDR uses a thermostable DNA ligase that ligates two adjacent oligonucleotides annealed to a complementary target only if the nucleotides are perfectly matched at the junction (1, 2). This method has subsequently been used in detecting mutations, insertions, and deletions in cancer genes (8–10, 21, 22). More recently this assay has been adapted for the detection of bacterial pathogens in blood cultures (34) and of West Nile virus in serum and mosquito pools (38). Since even a single base mismatch at the ligation junction prevents successful ligation, the technique is highly specific (21). Furthermore, such assays are ideal for multiplexing, since several primer sets can ligate along a DNA template without the interference encountered in purely polymerase-based assays (10, 22).

The LDR primers are designed to produce ligation products that are fluorescently labeled at their 3' ends and have zip-code complements (complementary to zip-code addresses in a universal array) appended to their 5' ends. The specific zip-code address spotted onto the array hybridizes only to the complementary sequence included on the LDR product (9, 10). A schematic representation of the assay is shown in Fig. 1. The

universal array is a powerful technique that permits the simultaneous detection of a large number of genes or gene products, making it ideal for use in multiplex, high-throughput assays.

Thus, the unique specificity and sensitivity of PCR-LDR coupled to the specificity of the universal array enables the detection and differentiation of all four serotypes of DENV in a single assay.

MATERIALS AND METHODS

Viral cultures and RNA from viruses and clinical samples. Viral culture supernatants were obtained from the Dengue Branch, Centers for Disease Control and Prevention (CDC), San Juan, Puerto Rico (n = 38). These included 10 isolates each of DENV-1 and -3 and 9 isolates each of DENV-2 and -4.

DENV strains selected from the global DENV panel (6) maintained at the Division of Vector-Borne Infectious Diseases (DVBID), CDC, Fort Collins, CO were used (Table 1). The panel consists of unique strains of DENV-1 to -4 representing the latest isolates of each serotype from different geographic regions. Isolates of other flaviviruses (Table 2) were kindly provided by Robert Lanciotti of the DVBID, CDC, the World Reference Center for Emerging Viruses and Arboviruses at the University of Texas Medical Branch in Galveston, M. Niedrig from the Robert Koch Institute, Berlin, Germany (European Network for Diagnostics of Imported Viral Diseases), and the New York City Department of Health (38).

A total of 350 clinical serum samples tested in this study were obtained from the repository at the Dengue Branch, CDC, San Juan, Puerto Rico where they were confirmed as either positive or negative for DENV by real-time PCR (5). These samples were collected from patients with suspected DF during epidemics and in the intraepidemic periods corresponding with the years when the different serotypes circulated in the region: 1994 to 1995 for DENV-1, 2002 to 2005 for DENV-2, 2000 to 2004 for DENV-3, and 1995 to 1998 for DENV-4. The serum specimens were collected from 25 different municipalities in Puerto Rico.

RNA extraction. Total RNA was extracted from 140 μl of human serum samples or virus-infected tissue culture supernatants by using the QIAamp viral RNA kit (Qiagen, Inc., Valencia, CA) according to the manufacturer's instructions. The RNA was extracted in 60 μl of elution buffer and was used immediately to synthesize the first-strand cDNA with the Superscript first-strand synthesis system for RT-PCR (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. Briefly, a master mix consisting of 0.1 mM deoxynucleoside triphosphates, 150 ng of random hexamers, 5 mM MgCl₂, 2 μl of 10× buffer, 10 mM dithiothreitol, 40 U of RNaseOUT recombinant RNase inhibitor, and 50 U of Superscript II reverse transcriptase enzyme was prepared. A 6-μl aliquot of extracted RNA was added to this mixture, and the sample was incubated at 42°C for 50 min, followed by 15 min at 70°C to terminate the reaction. An additional incubation of 20 min at 37°C with 1 μl of RNase H eliminated any residual RNA in the reaction product. The resultant cDNA was stored at −20°C for future use.

Primer design and assay development. PCR primers were designed using Oligo 6.0 software (Molecular Biology Insights, Cascade, CO) and were targeted to the capsid protein (C) and envelope protein (E) regions of the DENV genome. In the original version of the assay, nested forward and reverse primers were designed for the amplification of one region in gene C and two regions in gene E. Relatively conserved regions of the genome were chosen after alignment of known DENV sequences accessed from GenBank for primer design, but due to the extensive sequence variation in the DENV genome, different sets of primers had to be designed for each serotype, and degenerate bases were included to accommodate variation within each serotype. Primers had melting temperatures of approximately 72 to 75°C and were designed such that there were no more than three degenerate positions in each primer. Universal tail sequences were appended to the 5' ends of forward and reverse PCR primers to prevent the formation of primer dimers. A total of 23 PCR primers were designed for all four serotypes of DENV (see Tables S1 and S2 in the supplemental material).

LDR primers were designed at three positions within each of the gene E amplicons and two positions within the gene C amplicon. The upstream LDR primers had unique oligonucleotides (zip-code complements), 20 bases long, attached at the 5' end; the downstream primers had a fluorescent label, either 6-carboxyfluorescein (FAM) or cyanine 3 (Cy3), at the 3' end. A total of 128 LDR primers were designed (see Tables S3 and S4 in the supplemental material), with melting temperatures of 75 to 80°C; degenerate bases (no more than three in each primer) were introduced where required to account for sequence
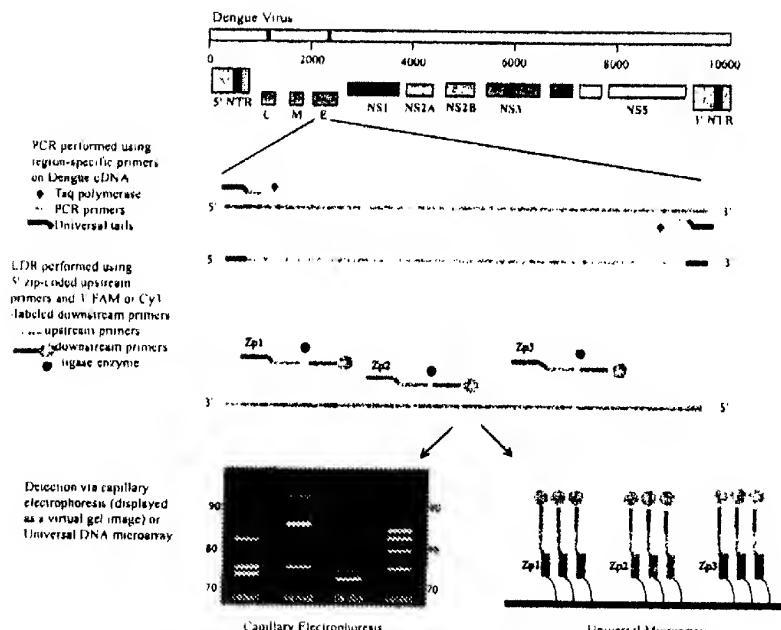
FIG. 1. Schematic of the PCR-LDR assay for the identification of DENV serotypes. Serotype-specific PCR primers amplify one region of gene C and one region of gene E (for clarity, only the gene E amplicon is shown). Within each PCR amplicon, LDR primers are designed to identify and differentiate the four different DENV serotypes. The LDR primers target two locations in gene C and three locations in gene E. The 5' upstream LDR primers bear zip-code complements, while the 3' downstream LDR primers have either a FAM or a Cy3 fluorescent label. Ligation of the LDR primers results in fluorescently labeled products of different lengths that are then detected either by CE or on a universal array. NTR, nontranslated region.

variations among different strains of the virus. Primers were obtained from Integrated DNA Technologies (Coralville, IA).

PCR-LDR assay and detection of products. Gene C and gene E were amplified using serotype-specific primers. The PCR mixture consisted of 10 mM Tris-HCl buffer containing 50 mM KCl (pH 8.0), 2.5 mM MgCl2, 0.8 mM deoxynucleoside triphosphates, and 1.25 U of AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA).

PCRs were optimized by performing serotype-specific uniplex reactions using DENV culture supernatants with 5 to 10 pmol of each primer per reaction. To this mixture, 1 µl of template cDNA was added to make a final volume of 25 µl. Amplification was performed using a GeneAmp 9700 thermocycler (Applied Biosystems, Foster City, CA). Initial denaturation of template DNA was achieved by heating at 95°C for 5 min. This was followed by 40 cycles of 30 s at 95°C, 30 s at 58°C, and 1 min at 72°C. A final extension step was conducted for 7 min at 72°C, followed by a termination step at 99°C for 30 min.

Multiplex PCRs were performed using the same method, except that all primers for all serotypes were used in a single reaction. Multiplex reactions were optimized by varying the concentrations of the primer (10 to 0.5 µM) and MgCl2 (1.5 to 2.5 mM). Ultimately, the optimum reaction conditions were found to be similar to those described above, except that the final concentration of each primer used was 2 µM and the MgCl2 concentration was 2.5 mM.

Ligation reactions were conducted in a solution (20 µl) containing LDR buffer (20 mM Tris-HCl buffer [pH 7.6], 100 mM KCl, 10 mM MgCl2), 1 mM NAD+, 1 mM dithiothreitol, 12.5 nM (250 fmol) each LDR primer, 2 µl of each PCR product, and 0.01 µM AK 16D ligase (expressed from Thermus species AK 16D) (42). Reaction mixtures were initially heated at 94°C for 1 min, followed by 30 thermal cycles at 94°C for 30 s (denaturation) and 64°C for 4 min (annealing/ligation). The ligation products were analyzed by two methods, capillary electrophoresis (CE) and a universal zip-code array.

CE. A 0.5-µl aliquot of each LDR product was added to 9.2 µl of Hi-Di formamide and 0.3 µl of a LIZ 500 DNA size standard (Applied Biosystems, Foster City, CA). The samples were denatured by heating to 95°C for 3 min and were cooled rapidly to 4°C before being loaded onto the ABI 3730 DNA analyzer for CE. The data generated were analyzed using Gene Mapper software (version 3.5; Applied Biosystems, Foster City, CA). The fragment size and peak area data of the different ligation products were exported and used to generate a 2-dimensional virtual gel image using the Gelvender software program (34). All reactions were performed in duplicate, and the experiment was run twice.

Universal zip-code array. Unique 20-bp oligonucleotides (zip-code addresses) were double spotted onto polymer-coated slides (CodeLink slides, GE Healthcare, Piscataway, NJ). Zip-code addresses for spotting were prepared in 50 mM sodium phosphate (pH 8.5) at a final concentration of 25 µM in a 384-well plate. A fiducial oligonucleotide (1 µM) was added to the printing mix in each well and cospotted with each zip-code address. Arrays were printed using a QArrayMini robotic array printer (Genetix, Boston, MA) at 10°C and 50 to 60% humidity. Printed slides were incubated in a saturated NaCl chamber overnight and then treated with a blocking solution (0.1 M Tris, 50 mM ethanolamine [pH 9.0]) to block residual reactive carboxyl groups. The slides were washed with 4× standard sodium citrate (SSC) buffer (20× SSC is 3 M sodium chloride and 0.3 M sodium citrate, pH 7.0) and 0.1% sodium dodecyl sulfate (SDS) and were dried by spinning. A 6-carboxy-X-rhodamine-labeled fiducial complement included in the hybridization mixture served as an internal positive control to determine the position and quality of each address. Printed slides were randomly selected for quality control by hybridizing fluorescent-labeled zip-code complements; a batch of slides was used only if the quality control produced a specific fluorescent signal in the absence of extraneous signals on adjacent addresses. Printed slides were stored in a desiccator at room temperature until use.

Ligation products were diluted in a hybridization buffer (5× SSC buffer)

TABLE 1. Global DENV panel used in assay validation

Sample ID	Origin	Yr of isolation	Source	Serotype	Genotype
276RKI	India	1997	Human	DENV-1	Genotype I
498 RKI	Thailand	1998	Human	DENV-1	Genotype I
1266	Indonesia	1978	Mosquito	DENV-1	Genotype II
12150	Philippines	1984	Human	DENV-1	Genotype II
228690	Jamaica	1977	Human	DENV-1	Genotype III
BC89/94	Costa Rica	1994	Human	DENV-1	Genotype III
10674	Dakar	1970	Human	DENV-2	Sylvatic genotype
BC102/94	Saudi Arabia	1994	Human	DENV-2	Cosmopolitan/genotype III
P8-1407MS	Malaysia	1968	Human	DENV-2	Cosmopolitan/genotype III
S-40921	Myanmar	1976	Human	DENV-2	Asian genotype I
BC27/96	Vietnam	1995	Unknown	DENV-2	Asian genotype II
BC171/96	Philippines	1996	Human	DENV-2	Asian genotype II
BC100/98	Bolivia	1998	Unknown	DENV-2	American/Asian
BC141/96	Puerto Rico	1994	Unknown	DENV-2	American/Asian
S-14635	Tonga	1974	Unknown	DENV-2	American
BC182/96	Philippines	1997	Human	DENV-3	Genotype I
BC14/97	Malaysia	1997	Human	DENV-3	Genotype I
MK-594-87	Thailand	1987	Unknown	DENV-3	Genotype II
S-40580	Myanmar	1976	Unknown	DENV 3	Genotype II
BC188/97	Mexico	1997	Human	DENV-3	Genotype III
271242	Sri Lanka	1991	Human	DENV-3	Genotype III
BC123/97	Malaysia	Unknown	Monkey	DENV-4	Sylvatic genotype
D85-019	Thailand	1985	Human	DENV-4	Genotype I
BC13/97	Malaysia	1997	Human	DENV-4	Genotype I
BC287/97	Mexico	1997	Human	DENV-4	Genotype II
BC258/97	Puerto Rico	1994	Human	DENV-4	Genotype II

containing 0.1% SDS, 0.1 mg/ml salmon sperm DNA (Fisher Scientific), and 5 nM fiducial complement in a total volume of 30 µl. A Pro-Plate multiarray slide chamber (Grace Bio-Labs, Bend, OR) was attached to universal array slides, and the entire amount of the hybridization mixture was added to the chambers. Hybridization was carried out at 60°C for 2 h in the dark in a hybridization oven (Lab-Line; VWR, West Chester, PA). Following hybridization, the slides were rinsed with 5× SSC and washed with 1× SSC–0.1% SDS at 60°C for 15 min. After two more wash steps of 1 min each with 0.2× SSC and 0.1× SSC, respectively at room temperature the slides were spin-dried and scanned using a Pro-Scan array (Perkin-Elmer, Wellesley, MA). Positive signals were detected and quantified with ScanArray Express (version 3.0; Perkin-Elmer, Wellesley, MA) and were manually inspected when necessary. The signal intensity data obtained were transferred as text files, and only those addresses where the signal intensity was >10-fold higher than background were considered positive. Reactions were performed twice and identification confirmed for both experiments.

LOD of the assay. To determine the limit of detection (LOD), 10-fold serial dilutions of viral culture stocks were prepared for all four DENV serotypes from standard stock cultures, with starting concentrations of 2,500,000 PFU/ml for DENV-1 (strain Hawaii), 290,000 PFU/ml for DENV-2 (strain New Guinea C), 1,200,000 PFU/ml for DENV-3 (strain Philippines H87), and 10,000,000 PFU/ml for DENV-4 (strain Philippines H241). The concentration ranges tested for the

TABLE 2. Details of other flavivirus strains tested

Flavivirus	Strain	Origin/yr
St. Louis encephalitis virus	MSI-7	Mississippi/1977
Murray Valley fever virus	OR2	Victoria, Australia/1951
Kunjin virus	MRM 16	Australia/1960
Powassan encephalitis virus	M11665	Ontario, Canada/1965
Yellow fever virus	17D	Ghana/1927
Japanese encephalitis virus	SA14-14-2	China/1954
West Nile virus	Unknown[a]	New York City/1999, 2000

[a] Unidentified strain collected from mosquito pools (38).

different serotypes were as follows: 2.5 × 10⁶ to 0.0025 PFU/ml (1.75 × 10⁵ to 1.75 × 10⁻⁵ PFU/reaction) for DENV-1, 2.9 × 10⁵ to 0.0029 PFU/ml (2 × 10⁴ to 2.03 × 10⁻⁵ PFU/reaction) for DENV-2, 1.2 × 10⁶ to 0.0012 PFU/ml (8.4 × 10⁴ to 0.84 × 10⁻⁵ PFU/reaction) for DENV-3, and 1 × 10⁷ to 0.01 PFU/ml (7 × 10⁵ to 0.7 × 10⁻⁶ PFU/reaction) for DENV-4. Dilutions were prepared in Dulbecco's minimum essential medium supplemented with 10% fetal bovine serum (Invitrogen, Carlsbad, CA). RNA was extracted from 140 µl of each dilution, and RT-PCR-LDR with a universal array was used to determine the lower limit of detection.

RESULTS

Assay design and initial optimization. Initial optimization and validation of the multiplex PCR-LDR assay for the detection and identification of DENV serotypes were performed on 38 culture supernatants from stock virus cultures. Products of PCR amplification were analyzed by electrophoresis on a 2% agarose gel. Both uniplex (each serotype amplified separately for each gene target) and multiplex (all serotypes amplified together for each gene target) reactions were individually validated, and PCR products were observed at 300 bp and 400 bp, respectively, for gene C and gene E (data not shown). LDRs were then performed on the PCR products, and the LDR products were analyzed by CE. The assay was able to successfully identify 37 out of the 38 samples and to type them correctly compared to standard detection by real-time PCR at the CDC in Puerto Rico (data not shown). The sample that failed identification was found to have insufficient nucleic acid. Our initial optimization and validation experiments for the detection of DENV serotypes were carried out successfully by analyzing LDR products from two amplicons (one each in gene E and gene C) (data not shown). Analysis of the third amplicon did not provide any additional information for the detection or differentiation of DENV serotypes. All further experiments

FIG 2. Reconstructed CE images of representative DENV isolates. Ligation products were labeled with FAM, separated by CE on a 3730 DNA analyzer, and sized using a LIZ-labeled internal standard. Each serotype produces a distinct pattern depending on the number and sizes of the ligation products obtained. The approximate sizes of the ligation products, which ranged from 73 to 92 bases, are indicated on each side. The CE data shown are reconstructed images produced by Gelrender software (34). Negative samples ($n = 7$) and samples from other viruses yielded no observable products. For other flaviviruses, lanes are as follows: 1, West Nile virus; 2, St. Louis encephalitis virus; 3, Japanese encephalitis virus; 4, Kunjin virus; 5, Murray Valley virus; 6, Powassan virus, 7, yellow fever virus.

were therefore performed using primers for only two amplicons: 46 PCR primers and 75 LDR primers in all (see Tables S1 to S4 in the supplemental material).

Assessment of the multiplex capability of the PCR-LDR assay. A multiplex PCR amplification of both target regions used all 46 PCR primers in a single reaction. Subsequently, amplicons were subjected to a multiplex ligation reaction using the 75 primers for all four DENV serotypes in a single reaction. For the initial validation, LDR products were analyzed by CE. Results from representative samples are shown in Fig. 2. The sizes and number of LDR products at a single position may differ depending on the target viral sequence in that particular ligation position. Table 3 shows the observed lengths of the LDR products for each serotype, which differed by 1 to 3 bases from those expected. A given DENV serotype can produce four or five different ligation products of approximately 73 to 92 bases. Due to variations in product length and the comigration of products of similar length, CE is not an optimal method for the interpretation of the PCR-LDR assay (Table 3 and Fig. 2). In a similar study, we have observed that LDR products of identical lengths but different sequences (due to the use of degenerate oligonucleotides) can migrate at separate positions in CE, resulting in broadened peaks that may overlap with nearby peaks (38). Other studies have also observed that it is difficult to predict CE results based on the lengths of the primers, because the migration pattern cannot be deduced from the length of the product (28). Therefore, after successful optimization of the PCR-LDR, the universal array was used for the analysis of LDR products.

TABLE 3 Comparison of the expected sizes of the ligation products and the sizes observed by CE analysis

Serotype	Target detected	Sizes of LDR products (bp)	
		Predicted	Observed
DENV-1	Gene C	76, 83	74.5, 76–76.2, 82.3 82.6
	Gene E	76, 78, 88	73.6–74, 75.9–76.2, 88.6–88.8
DENV-2	Gene C	76, 86	76.2–76.4, 85.9–86.0
	Gene E	85, 86, 92	82.4–82.6, 84.7–84.8, 91.7–91.9
DENV-3	Gene C	74, 74	73.7–73.9
	Gene E	75, 78, 78	75.2–75.4, 77.4–77.6
DENV-4	Gene C	79, 84	79.8–80.0, 81.3–81.4, 83.8–84.1
	Gene E	76, 83, 86	75.5–75.7, 82.6–82.8, 84.0–84.3

The detection of LDR products on the universal array is independent of their size. The serotype-specific products hybridize to unique complementary zip-code addresses on the array. The LDR products are detected by fluorescent signals appended to their 3' ends. A typical universal array layout is shown in Fig. 3A. All zip-code addresses were spotted in duplicate. Two positive signals were required, one from each amplicon or both from a single amplicon, for identification (Fig. 3B). Two additional zip-code addresses were spotted for DENV-4 due to the use of additional LDR primers designed for the detection of gene E; however, for any given DENV-4 serotype, only five positive signals were observed on the array. Two addresses on the array produced positive signals for both DENV-1 and DENV-3 due to similarities in the primer sequences. This did not, however, interfere with the identification of the two serotypes, because the other addresses clearly distinguished them (Fig. 3B).

Identification and typing of DENV from clinical samples. Serum specimens from 161 cases of DF confirmed at the Dengue Branch, CDC, Puerto Rico, were analyzed by the PCR-LDR assay. The results are shown in Table 4. It was possible to identify 159 of the 161 positive samples correctly (sensitivity, 98.7%). Serotype identification was concordant for all 159 samples in which DENV was detected. The two samples that tested positive at the CDC but could not be detected by PCR-LDR were archived samples; material could have been lost due to prolonged storage or multiple freeze-thaw cycles. Specificity was evaluated using a panel of 189 serum samples that tested negative for the presence of DENV. Three of the negative samples were identified as DENV-2 by our assay (specificity, 98.4%). Two of these samples came from patients presenting with acute symptoms of DF but were found to be DENV negative when tested by RT-PCR and an enzyme-linked immunosorbent assay at the CDC; no second, paired serum sample, which might have been tested for seroconversion, was ever received from either patient. No clinical data were available for the third patient, but the sample was found to be negative upon repeated analysis by real-time PCR.

Performance of the PCR-LDR against the global DENV panel and other flaviviruses. A global DENV panel consisting of six strains each of DENV-1 and -3, nine strains of DENV-2, and five strains of DENV-4 was tested using PCR-LDR with the universal array for detection and serotype identification
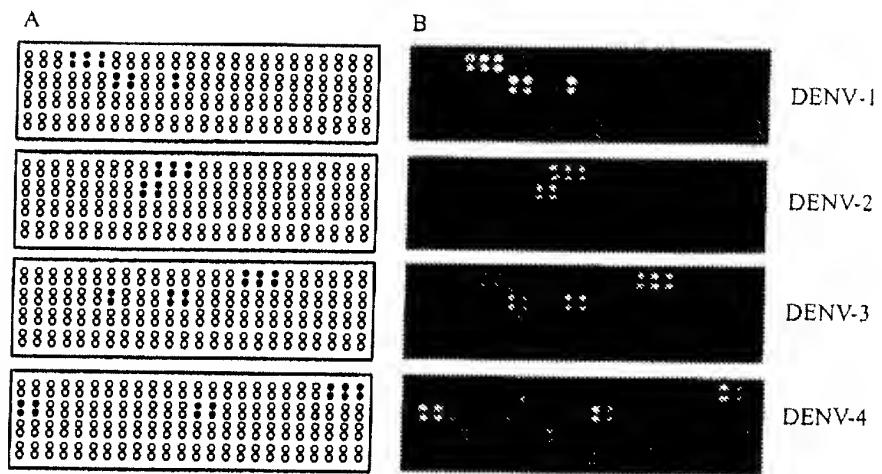
FIG. 3 Detection of DENV by the universal array. (A) Universal microarray layout of expected positive signals (filled circles) for each DENV serotype Zip-codes were spotted in duplicate. Open circles indicate addresses available for the detection of other viruses as the assay is further developed (B) Representative universal array detection of DENV-1 to -4 using the RT-PCR-LDR assay. Each serotype generates at least four or five unique signals that permit identification. DENV-1 and DENV-3 produce two common signals, as described in the text

The strains belonged to different genotypes and came from different geographic regions (Table 1) (6). All except one of the strains (strain identification [ID] 10674; Dakar, 1970) were detected and correctly serotyped by PCR-LDR Sequence analysis of gene E and gene C confirmed that the isolate had significant genetic variation at the PCR primer binding sites. Two additional PCR primers were subsequently designed and added to the PCR primer mix in order to accommodate the sequence difference and permit the amplification of this isolate (see Tables S1 and S2 in the supplemental material).

A panel of seven other flaviviruses (West Nile virus, Kunjin virus, Powassan virus, Japanese encephalitis virus, St. Louis encephalitis virus, Murray Valley encephalitis virus, and yellow fever virus) was tested with the assay, and CE results are shown in Fig 2. The PCR-LDR primers were specific for DENV serotypes, and no signals were detected for any of the other flaviviruses by use of CE or the universal array

Determination of LOD. Using serial dilutions of the virus stock cultures for RNA extraction and subsequent PCR-LDR, detection limits were found to be 0.017, 0.004, 0.8, and 0.7

TABLE 4 Detection and serotyping of DENV from clinical serum samples using a multiplex RT-PCR-LDR-universal array

CDC serotype determination	No of samples tested	No of samples with the following serotype by the RT-PCR-LDR-universal array				
		DENV-1	DENV-2	DENV-3	DENV-4	Negative
DENV-1	20	20	0	0	0	0
DENV-2	62	0	62	0	0	0
DENV-3	59	0	0	57	0	2
DENV-4	20	0	0	0	20	0
Negative	189	0	3	0	0	186

equivalent PFU of virus for DENV-1, DENV-2, DENV-3, and DENV-4, respectively. Figure 4 shows the universal array data for the DENV-1 detection limit. Detection limits were found to be the same by using CE or the universal array and were calculated by taking the dilution factors into account; i e., RNA was extracted from 140 μl of each dilution and was subsequently eluted out in 60 μl of buffer, from which 6 μl was used for cDNA synthesis.

DISCUSSION

In this study, we report the development and evaluation of a nucleic acid detection assay, based on RT-PCR followed by LDR with specific primers, for the simultaneous identification of all four serotypes of DENV in a single reaction.

We used the PCR-LDR approach to overcome the high degree of sequence variation between DENV serotypes. The wide distribution of DENV serotypes in the tropics and subtropics allows for considerable genotypic heterogeneity among the circulating serotypes and has also been linked to the virulence of the circulating strains (5, 12, 17, 23, 30). The four different serotypes of DENV are thought to be more dissimilar than different "species" of *Flavivirus* (17). Such viral genetic heterogeneity has implications for the design of molecular assays. False-negative PCR results due to sequence mismatches between DENV RNA and the primers used in real-time PCR assays have been reported, necessitating assay revisions (6, 16, 36). In the design of our assay, we used several primer pairs with degeneracies to accommodate possible failures of amplification due to sequence mismatches The envelope (E) and capsid (C) genes were chosen as targets, because they were found to have sufficient variation to be useful for

FIG. 4. LOD of the PCR-LDR assay. Microarray images show the LOD for DENV-1. The number of PFU used in each PCR is given. Fluorescent signals ≥10-fold higher than the background negative signals were deemed positive.

serotype differentiation. However, due to the inherent heterogeneity in the E region, a large number of primers had to be used to account for all anticipated sequence variations. Primers were designed for two regions in order to provide a certain amount of redundancy that would circumvent any dropouts in PCR amplification due to sequence variation.

A multiplex PCR-LDR assay was optimized such that all primers could be used in a single reaction for amplification of all four DENV serotypes. This approach was found to be >98% sensitive and specific in the detection and serotype identification of DENV from 350 archived acute-phase serum samples and to compare favorably to techniques described previously (11, 24, 40). The assay has been developed as a prototype for the multiplex detection of RNA viruses and will be expanded for the detection of other hemorrhagic fever viruses. Therefore, a multiplex format was adopted to define the feasibility of the assay, and it was possible to multiplex 46 PCR primers and 75 LDR primers in a single reaction, with significant signal intensity in the CE and microarray formats.

The primers were very specific for DENV serotypes and did not cross-react with any of the other flaviviruses tested. Only 3 of 189 samples were positive by PCR-LDR but negative by the real-time PCR assay performed at the CDC. Upon analysis of clinical findings, it was discovered that the serum samples from two of these patients were collected within 1 to 5 days of the onset of symptoms, suggesting that they may have had acute infections. Seroconversion could not be demonstrated, because convalescent-phase serum samples were never obtained. It is therefore possible that these patients indeed had viremia that could not be detected by real-time PCR. No clinical history was available for the third patient whose sample was retrieved from the CDC repository archives.

Current reports suggest that several lineages of DENV-2 and DENV-4 circulate in Puerto Rico and that DENV strains exhibit complex patterns of lineage turnover and extensions (3, 4). Thus, it is likely that several lineages of each DENV serotype could have been circulating within the time period encompassing the collection of specimens included in the panel

of clinical samples used for this study, which were detected by our assay with high sensitivity and specificity.

In order to verify the ability of the PCR-LDR assay to detect different genotypes of DENV, we used a global DENV panel. Our assay was able to detect all but one isolate (strain ID 10674), a sylvatic strain from Africa (G.-J. J. Chang, personal communication). The analysis of complete genome sequences suggests that sylvatic DENV-2 isolates are evolutionarily distinct from endemic DENV-2 isolates and supports the classification of DENV-2 into two discrete ecotypes (4, 43). In another study, real-time PCR primers for DENV targeted to the C-PrM region had to be modified to detect variant sylvatic virus (6; G.-J. J. Chang, personal communication). The sequence of sylvatic isolate 10674 was not available when our primers were initially designed; subsequently, two new PCR primers were added, and it was possible to amplify this isolate. The addition of two new primers did not adversely affect the efficiency of the PCR, indicating the flexibility of our assay in the identification of new genotypes. The evolutionary genetics of DENV suggest that its population is becoming increasingly diverse, and the presence of sylvatic reservoirs of the virus may allow the introduction and emergence of sylvatic virus in human populations. It has also been suggested that these conditions could lead to the development of new pathogenic DENV strains (17). With the PCR-LDR strategy, it would be possible to incorporate new primers for the detection of these variant genotypes as they evolve.

The detection limit of our assay ranged from 0.004 to 0.7 equivalent PFU/reaction. The lower LOD differed for the different serotypes; the assay was 100 times more sensitive for DENV-2 and DENV-1 (LOD, 0.004 and 0.017 equivalent PFU, respectively) than for DENV-3 and DENV-4 (LOD, 0.8 and 0.7 equivalent PFU, respectively). Variable LOD have been reported for DENV serotypes in several different studies. Johnson et al. have reported an assay 100 times more sensitive for DENV-2, -3, and -4 (0.0016 to 0.008 equivalent PFU) than for DENV-1 (0.5 equivalent PFU) and postulated that the difference could be due to differences in the proportion of

noninfectious RNA transcripts to infectious particles (PFU) between DENV-1 and the other serotypes (20). In a comparable study, Lai et al have reported variable LOD of 0.1 PFU for DENV-1 and -2, 1 PFU for DENV-3, and 0.01 PFU for DENV-4 (26) In yet another real-time PCR-based assay evaluated recently, serotype-specific primers were 10 times more sensitive for DENV-2, -3, and -4 (LOD, 0.1 PFU) than for DENV-1 (LOD, 1 PFU) (33). The LOD observed in the present study is comparable to those for the other techniques reported (20, 26, 40).

The use of zip-code addresses for spotting the array and its unique potential to recognize pathogen-specific zip-code complements appended to the LDR primers obviate the use of actual genetic sequence for pathogen detection. This is especially useful for multiplexing: a large number of genomic targets can be detected using this array, since the zip-code addresses are synthetic oligonucleotides. Additionally, the same array can be used for the simultaneous detection of different organisms, since positive hybridization is dependent on the chemistry of the synthesized zip-code oligonucleotides spotted onto the array and their complements appended to the primers.

The current study was undertaken to prove the feasibility of using the PCR-LDR technique for the multiplexed detection of RNA viruses. Since viral infections often present with nonspecific clinical symptoms, and arboviruses have similar geographical distributions, a multiplexed approach to the detection of several arboviruses in a single sample could be extremely beneficial. In addition to serotype identification, a critical element of epidemic control measures is the early identification of emerging genotypes and the replacement of a genotype(s) in a given geographic region (37) Since some DENV serotypes have more than four genotypes, PCR-LDR may be useful in genotype detection and differentiation without the involvement of nucleotide sequence analysis and phylogenetic studies.

We envision extending the scope of the assay to permit the multiplexed identification of a panel of hemorrhagic fever viruses. Our group has already reported the use of the PCR-LDR assay for the identification of a panel of 20 different bacteria in blood cultures (34) and has developed an assay for the detection of West Nile virus (38). The DENV detection assay has been developed using similar principles, and the technique is currently being developed for use in the simultaneous identification of hemorrhagic fever viruses. Ultimately the technique may prove useful in a comprehensive assay for blood-borne infectious agents.

ACKNOWLEDGMENTS

REFERENCES

1. Barany, F. 1991 Genetic disease detection and DNA amplification using cloned thermostable ligase Proc. Natl. Acad. Sci. USA 88:189–193

2. Barany, F. 1991 The ligase chain reaction in a PCR world. PCR Methods Appl. 1:5–16

3. Bennett, S. N., E. C. Holmes, M. Chirivella, D. M. Rodriguez, M. Beltran, V. Vorndam, D. J. Gubler, and W. O. McMillan. 2006. Molecular evolution of dengue 2 virus in Puerto Rico: positive selection in the viral envelope accompanies clade reintroduction. J. Gen. Virol. 87:885–893.

4. Bennett, S. N., E. C. Holmes, M. Chirivella, D. M. Rodriguez, M. Beltran, V. Vorndam, D. J. Gubler, and W. O. McMillan. 2003. Selection-driven evolution of emergent dengue virus Mol. Biol. Evol. 20:1650–1658.

5. Bray, M., R. Men, I. Tokimatsu, and C. J. Lai, 1998. Genetic determinants responsible for acquisition of dengue type 2 virus mouse neurovirulence J. Virol. 72:1647–1651

6. Chien, L. J., T. L. Liao, P. Y. Shu, J. H. Huang, D. J. Gubler, and C. J. Chang. 2006 Development of real-time reverse transcriptase PCR assays to detect and serotype dengue viruses. J. Clin. Microbiol. 44:1295–1304

7. Dupuis, A. P., II, P. P. Marra, R. Reitsma, M. J. Jones, K. L. Louie, and L. D. Kramer. 2005. Serologic evidence for West Nile virus transmission in Puerto Rico and Cuba. Am. J. Trop. Med. Hyg. 73:474–476.

8. Favis, R., and F. Barany, 2000 Mutation detection in K-ras, BRCA1, BRCA2, and p53 using PCR/LDR and a universal DNA microarray Ann N Y Acad Sci 906:39–43.

9. Favis, R., J. P. Day, N. P. Gerry, C. Phelan, S. Narod, and F. Barany, 2000 Universal DNA array detection of small insertions and deletions in BRCA1 and BRCA2. Nat. Biotechnol. 18:561–564

10. Gerry, N. P., N. E. Witowski, J. Day, R. P. Hammer, G. Barany, and F. Barany. 1999 Universal DNA microarray method for multiplex detection of low abundance point mutations J. Mol. Biol. 292:251–262

11. Gomes, A. L., A. M. Silva, M. T. Cordeiro, G. F. Guimaraes, E. T. Marques, Jr., and F. G. Abath, 2007 Single-tube nested PCR using immobilized internal primers for the identification of dengue virus serotypes J. Virol. Methods 145:76–79

12. Gualano, R. C., M. J. Pryor, M. R. Cauchi, P. J. Wright, and A. O Davidson. 1998 Identification of a major determinant of mouse neurovirulence of dengue virus type 2 using stably cloned genomic-length cDNA J. Gen. Virol. 79:437–446

13. Gubler, D. J. 1998 Dengue and dengue hemorrhagic fever Clin. Microbiol Rev. 11:480–496

14. Gubler, D. J. 1998. The global pandemic of dengue/dengue haemorrhagic fever: current status and prospects for the future Ann Acad Med Singapore 27:227–234

15. Guzmán, M. G., and G. Kouri. 2004 Dengue diagnosis, advances and challenges Int. J. Infect. Dis 8:69–80

16. Harris, E., T. G. Roberts, L. Smith, J. Selle, L. D. Kramer, S. Valle, E. Sandoval, and A. Balmaseda, 1998. Typing of dengue viruses in clinical specimens and mosquitoes by single-tube multiplex reverse transcriptase PCR. J. Clin. Microbiol. 36:2634–2639.

17. Holmes, E. C., and S. S. Burch. 2000. The causes and consequences of genetic variation in dengue virus. Trends Microbiol. 8:74–77

18. Ito, M., T. Takasaki, K. Yamada, R. Nerome, S. Tajima, and I. Kurane. 2004 Development and evaluation of fluorogenic TaqMan reverse transcriptase PCR assays for detection of dengue virus types 1 to 4 J. Clin. Microbiol. 42:5935–5937.

19. Ito, M., K. Yamada, T. Takasaki, B. Pandey, R. Nerome, S. Tajima, K. Morita, and I. Kurane. 2007 Phylogenetic analysis of dengue viruses isolated from imported dengue patients: possible aid for determining the countries where infections occurred. J. Travel Med. 14:233–244.

20. Johnson, B. W., B. J. Russell, and R. S. Lanciotti. 2005 Serotype-specific detection of dengue viruses in a fourplex real-time reverse transcriptase PCR assay. J. Clin. Microbiol 43:4977–4983.

21. Khanna, M., W. Cao, M. Zirvi, P. Paty, and F. Barany. 1999 Ligase detection reaction for identification of low abundance mutations Clin. Biochem 32:287–290.

22. Khanna, M., P. Park, M. Zirvi, W. Cao, A. Picun, J. Day, P. Paty, and F. Barany. 1999 Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumors. Oncogene 18:27–38.

23. Kinney, R. M., S. Butrapet, G. J. Chang, K. R. Tsuchiya, J. T. Roehrig, N. Bhamarapravati, and D. J. Gubler. 1997 Construction of infectious cDNA clones for dengue 2 virus strain 16681 and its attenuated vaccine derivative strain PDK-53 Virology 230:300–308

24. Klungthong, C., R. V. Gibbons, B. Thaisomboonsuk, A. Nisalak, S. Kalayanarooj, V. Thirawuth, N. Nutkumhang, M. P. Mammen, Jr., and R G Jarman. 2007 Dengue virus detection using whole blood for reverse transcriptase PCR and virus isolation J Clin Microbiol 45:2480–2485

25. Komar, N., and G. G. Clark. 2006 West Nile virus activity in Latin America and the Caribbean Rev. Panam Salud Pública 19:112–117

26. Lai, Y. L., Y. K. Chung, H. C. Tan, H. F. Yap, G. Yap, E. E. Ooi, and L. C. Ng. 2007. Cost-effective real-time reverse transcriptase PCR to screen for dengue virus followed by rapid single-tube multiplex RT-PCR for serotyping of the virus J. Clin. Microbiol 45:935–941

27. Lanciotti, R. S., C. H. Calisher, D. J. Gubler, G. J. Chang, and A. V. Vorndam. 1992. Rapid detection and typing of dengue viruses from clinical

samples by using reverse transcriptase-polymerase chain reaction. J. Clin. Microbiol. 30:545-551.

28. Larsen, L. A., C. Jespergaard, and P. S. Andersen. 2007. Single-strand conformation polymorphism analysis using capillary array electrophoresis for large-scale mutation detection. Nat. Protoc. 2:1458-1466.

29. Laue, T., P. Emmerich, and H. Schmitz. 1999. Detection of dengue virus RNA in patients after primary or secondary dengue infection by using the TaqMan automated amplification system. J. Clin. Microbiol. 37:2543-2547.

30. Leitmeyer, K. C., D. W. Vaughn, D. M. Watts, R. Salas, I. Villalobos de Chacun, C. Ramos, and R. Rico-Hesse. 1999. Dengue virus structural differences that correlate with pathogenesis. J. Virol. 73:4738-4747.

31. Mackenzie, J. S., D. J. Gubler, and L. R. Petersen. 2004. Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. Nat. Med. 10:S98-S109.

32. Morens, D. M., and A. S. Fauci. 2008. Dengue and hemorrhagic fever: a potential threat to public health in the United States. JAMA 299:214-216.

33. Parida, M., K. Horioke, H. Ishida, P. K. Dash, P. Saxena, A. M. Jana, M. A. Islam, S. Inoue, N. Hosaka, and K. Morita. 2005. Rapid detection and differentiation of dengue virus serotypes by a real-time reverse transcription-loop-mediated isothermal amplification assay. J. Clin. Microbiol. 43:2895-2903.

34. Pingle, M. R., K. Granger, P. Feinberg, R. Shatsky, B. Sterling, M. Rundell, E. Spitzer, D. Larone, L. Golightly, and F. Barany. 2007. Multiplexed identification of blood-borne bacterial pathogens by use of a novel 16S rRNA gene PCR-ligase detection reaction-capillary electrophoresis assay. J. Clin. Microbiol. 45:1927-1935.

35. Porterfield, J. S. 1986. Antibody-dependent enhancement of viral infectivity. Adv. Virus Res. 31:335-355.

36. Reynes, J. M., S. Ong, C. Mey, C. Ngan, S. Hoyer, and A. A. Sall. 2003. Improved molecular detection of dengue virus serotype 1 variants. J. Clin. Microbiol. 41:3864-3867.

37. Rico-Hesse, R., L. M. Harrison, R. A. Salas, D. Tovar, A. Nisalak, C. Ramos, J. Boshell, M. T. de Mesa, R. M. Nogueira, and A. T. da Rosa. 1997. Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. Virology 230:244-251.

38. Rondini, S., M. R. Pingle, S. Das, R. Tesh, M. S. Rundell, J. Hom, S. Stramer, K. Turner, S. N. Rossmann, R. Lanciotti, E. G. Spier, J. Muñoz-Jordán, D. Larone, E. Spitzer, F. Barany, and L. M. Golightly. 2008. Development of multiplex PCR-ligase detection reaction assay for detection of West Nile virus. J. Clin. Microbiol. 46:2269-2279.

39. Rothman, A. L. 2004. Dengue: defining protective versus pathologic immunity. J. Clin. Invest. 113:946-951.

40. Shu, P. Y., S. F. Chang, Y. C. Kun, Y. Y. Yueh, L. J. Chien, C. L. Sue, T. H. Lin, and J. H. Huang. 2003. Development of group- and serotype-specific one-step SYBR green I-based real-time reverse transcription-PCR assay for dengue virus. J. Clin. Microbiol. 41:2408-2416.

41. Tavakoli, N. P., E. H. Tobin, S. J. Wong, A. P. Dupuis II, B. Glasheen, L. D. Kramer, and K. A. Bernard. 2007. Identification of dengue virus in respiratory specimens from a patient who had recently traveled from a region where dengue virus infection is endemic. J. Clin. Microbiol. 45:1523-1527.

42. Tong, J., W. Cao, and F. Barany. 1999. Biochemical properties of a high fidelity DNA ligase from Thermus species AK16D. Nucleic Acids Res. 27:788-794.

43. Vasilakis, N., E. B. Fokam, C. T. Hanson, E. Weinberg, A. A. Sall, S. S. Whitehead, K. A. Hanley, and S. C. Weaver. 2008. Genetic and phenotypic characterization of sylvatic dengue virus type 2 strains. Virology 377:296-307.

44. Warrilow, D., J. A. Northill, A. Pyke, and G. A. Smith. 2002. Single rapid TaqMan fluorogenic probe based PCR assay that detects all four dengue serotypes. J. Med. Virol. 66:524-528.

45. World Health Organization. 2002. Dengue and dengue hemorrhagic fever. WHO fact sheet 117. World Health Organization, Geneva, Switzerland.

Exhibit 13:   Rondini et al, "Development of Multiplex PCR-Ligase Detection Reaction Assay for Detection of West Nile Virus," *J. Clin. Microbiol.* 46:2269–79 (2008)

Development of Multiplex PCR-Ligase Detection Reaction Assay for Detection of West Nile Virus[▽]

S. Rondini,[1] M. R. Pingle,[2] S. Das,[1] R. Tesh,[3] M. S. Rundell,[2] J. Hom,[4] S. Stramer,[5] K. Turner,[6]
S. N. Rossmann,[6] R. Lanciotti,[7] E. G. Spier,[8] J. Muñoz-Jordán,[9] D. Larone,[10] E. Spitzer,[11]
F. Barany,[2] and L. M. Golightly[1]*

Department of Medicine, Division of International Medicine and Infectious Diseases, Weill Medical College of Cornell University, New York, New York[1]; Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, New York[2]; Department of Pathology, University of Texas Medical Branch, Galveston, Texas[3]; New York City Department of Health and Mental Hygiene, Public Health Laboratory Virology & Immunology Division, New York, New York[4]; Scientific Support Office, American Red Cross Biomedical Services, Gaithersburg, Maryland[5]; Gulf Coast Regional Blood Center, Houston, Texas[6]; Centers for Disease Control and Prevention, Fort Collins, Colorado[7]; Applied Biosystems, Foster City, California[8]; Centers for Disease Control and Prevention, San Juan, Puerto Rico[9]; Department of Pathology, Weill Medical College of Cornell University, New York, New York[10]; and Department of Pathology, Stony Brook University Medical Center, Stony Brook, New York[11]

Received 4 December 2007/Returned for modification 16 January 2008/Accepted 1 May 2008

We have developed a novel multiplex reverse transcription-PCR ligase detection reaction (RT-PCR/LDR) assay for the detection of West Nile virus (WNV) in both clinical and mosquito pool samples. The method relies on the amplification of three different genomic regions, one in the coding sequence of nonstructural protein NS2a and two in nonstructural protein NS5, to minimize the risk of detection failure due to genetic variation. The sensitivity of the PCR is complemented by the high specificity of the LDR step, and the detection of the LDR products can be achieved with capillary electrophoresis (CE) or a universal DNA microarray. We evaluated the limit of detection by both one-step and two-step multiplex RT-PCR/LDR/CE approaches, which reached, respectively, 0.005 and 0.017 PFU. The assay demonstrated 99% sensitivity when mosquito pool samples were tested and 100% sensitivity with clinical samples when the one-step approach was used. The broad strain coverage was confirmed by testing 34 WNV isolates belonging to lineages 1 and 2, and the high specificity of the assay was determined by testing other flaviviruses, as well as negative mosquito pool and clinical samples. In summary, the multiplex RT-PCR/LDR assay could represent a valuable complement to WNV serological diagnosis, especially in early symptomatic patients. In addition, the multiplexing capacity of the technique, which can be coupled to universal DNA microarray detection, makes it an amenable tool to develop a more comprehensive assay for viral pathogens.

West Nile virus (WNV) was first isolated in 1937 from a patient in Uganda (32) and has since become the most widespread member of the Japanese encephalitis virus (JEV) complex (2). It made its first appearance in the western hemisphere in 1999, when 62 cases of WNV encephalitis and seven fatalities were reported in New York City (10, 22, 26). A surveillance program of mosquito pools was initiated by the New York City Department of Health and Mental Hygiene in 2000 (36, 37), and blood donor screening was introduced in the United States in June 2003 after the report of WNV transmission to 23 individuals from 16 viremic donors in the preceding year (8, 33). Since its introduction into the United States, the virus has rapidly spread and it is now endemic in 41 states and the District of Columbia (11).

WNV strains belong primarily to two different lineages, which exhibit considerable genomic diversity, thus posing challenges to their identification and detection in both clinical and environmental samples. Lineage 1 comprises strains from several continents, and it can be subdivided into at least three clades: clade 1a, which includes strains from Europe, Africa,

the United States and Israel; clade 1b, which is specific for Kunjin isolates from Australia; and clade 1c, which includes two strains from India. Lineage 2 comprises WNV strains from sub-Saharan regions of Africa and from Madagascar (6, 20). The genetic distance between lineage 2 isolates is particularly great (75.7 to 76.8%) compared to that observed in lineage 1 strains (95.2 to 99.9%) (20).

In addition to this inherent diversity, the spread of the various lineages of WNV and the isolation of new viral strains are constantly being recognized. A lineage 2 strain was recently reported in central Europe (2), and two new isolates from central Europe and southern Russia have recently been characterized which, based on their genetic distances, can be considered either separate lineages of WNV (lineages 3 and 4) or new members of the JEV group (1). In addition, phylogenetic analyses of North American isolates have shown an accumulation of mutations in the genome of prototype New York strain WN-NY99 leading to new genetic variants (5, 13, 20).

Screening of both mosquito pools and the blood supply is performed primarily by reverse transcription (RT)-PCR-based nucleic acid amplification testing (NAT) methods (31, 33). Despite the genomic diversity, a recent proficiency study revealed that less than 40% of the participating laboratories could detect lineage 2 strains (28). Molecular detection methods which are sensitive yet capable of detecting both lineages,

* Corresponding author. Mailing address: Weill Medical College of Cornell University, Division of International Medicine and ID, 1300 York Ave., Room A421, New York, NY 10021. Phone: (212) 746-6320. Fax: (212) 746-8675. E-mail: Lgolight@med.cornell.edu.
▽ Published ahead of print on 21 May 2008.
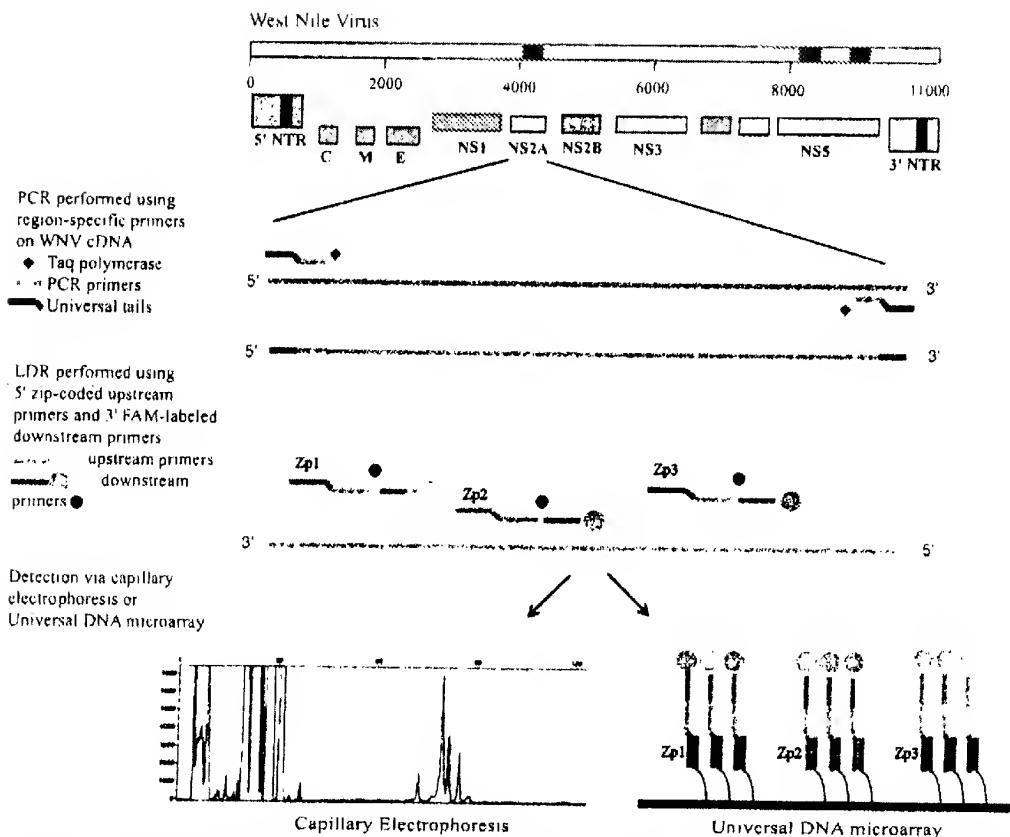
FIG. 1. Schematic of the multiplex RT-PCR/LDR assay for WNV. Multiple PCR primer pairs are designed to amplify three distinct regions of the WNV cDNA (in black). Each PCR primer contains between one and three degenerate positions to accommodate minor sequence variation at the primer binding sites. For simplicity, only one PCR amplicon is shown. The presence of each PCR amplicon is detected by LDR primer pairs specific for three regions within each amplicon. The 5' upstream LDR primers bear zip code complements, while the 3' LDR downstream primers bear a 6-carboxyfluorescein (FAM) fluorescent label. Ligation of an LDR primer pair results in fluorescently labeled products of different lengths that can be detected either by CE or by hybridization to a universal DNA microarray. NTR, nontranslated region.

as well as the numerous and emerging strains of WNV, are therefore needed (34).

In this report, we describe the development of a new, sensitive assay for the detection of both lineages of WNV, based on multiplex RT-PCR and the ligase detection reaction (LDR) (Fig. 1). LDR was originally developed for discriminating single-base mutations or polymorphisms (3, 4). The technique is amenable to multiplexing and has been used to detect mutations and single nucleotide polymorphisms in cancer genes (14–16, 18, 19) and recently to detect and identify bacteria in clinical blood samples (29).

We describe the validation of the technique for WNV detection and identification with mosquito pool samples, clinical isolates, and national, as well as international, strains. The high

sensitivity and broad strain coverage of the multiplex RT-PCR/LDR assay could render it a valuable complement to WNV serological diagnosis, especially in early symptomatic patients. In addition, the multiplexing capacity of the technique makes it amenable to the development of a more comprehensive assay for viral pathogens.

MATERIALS AND METHODS

TABLE 1. WNV strains used in this study

Designation	Strain	Origin/yr	Lineage-clade(s) (reference)
CAR67	ArB-310/67	Central African Republic/1967	1-b
NIG65	IbAn7019	Nigeria/1965	1-b
SEN79	ArD-27875	Senegal/1979	1-b
USA99a	31A	United States/1999	1-b
IND68	68856	India/1968	1-a
EGY51	Egypt101	Egypt/1951	1-a
ETH76	EthAn4766	Ethiopia/1976	1-a
AUS60	MRM16(Kunjin)	Australia/1960	KUN 1-b
AUS91	K6453(Kunjin)	Australia/1991	KUN 1-b
IND57	IG-15578	India/1957	IND 1-c, 5 (9)[a]
IND80	804994	Bangalore, India/1980	IND 1-c, 5 (9)[a]
SEN920	ArD-76104	Senegal/1990	2
CAR82	ArB3573/82	Central African Republic/1982	2
SA58	SAH-442	South Africa/1958	2
SA89	SPU116-89	South Africa/1989	2
MAD-88	ArMg-979	Madagascar/1988	2
MAD78	DakAnMg78	Madagascar/1978	2
CYP68	Q3574-5	Cyprus/1968	2
RabV	97-103	Czech Republic/1997	3 (1)[b]
WNV-Ug37	B 956 (WNFCG)	Uganda/1937	2
New York99	NY99	United States/1999	1
PaAn001	M12294 Fr00	France/2000	1
WNV 0043	WNV 0043	Israel/2000	1
MXH 442	TVP-9497	Sonora, Mexico/2004	1
JA 377-03	TVP 9115	Tabasco, Mexico/2003	1
TX 2676	TVP 10267	Texas/2006	1
AR 2771	TVP 8833	Quebec, Canada/2002	1
15476-04	TVP 9241	Nebraska/2003	1
IL-10861	TVP 9536	Illinois/2004	1
CT 8495-04	TVP 10030	Connecticut/2004	1
TX 5784	TVP10122	Texas/2006	1
101-2032	TVP 8852	Florida/2001	1
LA 02-01	TVP 9177	California/2004	1
GRLA 1260	TVP 9744	California/2003	1

[a] Classified also as a new distinct fifth lineage (clade 5; see Discussion).
[b] Classified either as a new third lineage or as a novel flavivirus within the JEV group.

For sensitivity testing, a WNV load panel containing plasma samples spiked with defined dilutions of NY99 virus was kindly provided by R. Lanciotti from the Centers for Disease Control and Prevention (CDC) Division of Vector-Borne Infectious Diseases in Fort Collins, CO (21). The panel titers ranged from 180,000 to 0.15 PFU/ml (quantified by standard plaque assay). Other flaviviruses used to determine the specificity of the assay were obtained from various sources. The St. Louis encephalitis, Murray Valley fever, Powassan encephalitis, and yellow fever viruses were obtained from R. Lanciotti (CDC, Fort Collins); JEV and tick-borne encephalitis virus were obtained from M. Niedrig (Robert Koch Institute, Berlin, Germany); and dengue virus serotypes I to IV were obtained from J. Muñoz (CDC Dengue Branch in San Juan, Puerto Rico) (Table 2).

Mosquito pools and clinical samples. Ninety-eight positive and 20 negative mosquito pool samples were obtained from the New York City Department of Health and Mental Hygiene (NYC DOHMH), which collects and tests trapped

mosquitoes from different areas in New York City as part of its WNV surveillance program (36, 37).

Plasma samples from 50 NAT-positive blood donors and 92 NAT-negative donors were tested. The positive plasma samples were provided by the Gulf Coast Regional Blood Center in Houston, TX; the negative samples were obtained from the CDC Dengue branch in Puerto Rico. Twenty additional plasma samples which tested positive for dengue virus and negative for WNV were also provided by the CDC in Puerto Rico. Additional samples for initial primer validation were obtained from S. Stramer at the American Red Cross National Testing and Reference Laboratories, Gaithersburg, MD.

RNA extraction. Samples from mosquito pool homogenates were clarified by centrifugation at 2,000 × g for 3 min. Viral RNA was extracted from the mosquito pool sample supernatants and clinical samples, as well as from viral seeds, with the QIAamp viral RNA mini kit (Qiagen, Valencia, CA) according to

TABLE 2. Other flaviviruses used in this study

Flavivirus	Strain	Origin/yr	Titer
Dengue virus serotypes I, II, III, IV	ND[a]	Puerto Rico/2004-2006	ND
St. Louis encephalitis virus	MSI-7	Mississippi/1977	3×10^8 PFU/ml
Murray Valley fever virus	OR2	Victoria, Australia/1951	ND
Powassan encephalitis virus	M11665	Ontario, Canada/1965	ND
Yellow fever virus	17D	Ghana/1927	5×10^5 PFU/ml
JEV	SA14-14-2	China/1954	2×10^7 copies/ml
Tick-borne encephalitis virus	K23	Reference 27	1×10^7 copies/ml

[a] ND, not determined.

TABLE 3. Primers used for PCR and RT-PCR

Target region	Forward primer sequence[a] (5'-3')	T_m (°C)[b]	Reverse primer sequence[a] (5'-3')	T_m (°C)	Amplicon length (bp)
1 (NS2a)	WNV_PCR_F5, <u>GCCAACTACCGCAACACA</u>GYTGGGCCTTMTGGTCGTGTTC	81–83	WNV_PCR_R7A, <u>CCAACTACCGCAACCA</u>CTTTGATGAGGCTTCCAACTCCAACCAT	81	540–580
	WNV_PCR_F6, <u>GCCAACTACCGCAACACT</u>GGCCACCCAGGAGGTCCTTC	83	WNV_PCR_R7B, <u>CCAACTACCGCAACCA</u>CCCTGATCAARCTGCCTATTCCRACCAT	81–83	
			WNV_PCR_R8A, <u>CCAACTACCGCAACCA</u>ATGAGGCAAGCTCCTTTCTTTTTTGC	79	
			WNV_PCR_R8B, <u>CCAACTACCGCAACCA</u>ARCAGACTKGCTCCTTTCTTYTTTGC	78–81	
2 (NS5)	WNV_PCR_F9A, <u>GCCAACTACCGCAACAC</u>GATGTGGAAGAGGCGGYTGGTGTTA	82–83	WNV_PCR_R11A, <u>CCAACTACCGCAACC</u>ACCCAGAAGCACCTGGCTWGTCAT	81	460–510
	WNV_PCR_F9B, <u>GCCAACTACCGCAACAC</u>GGTGTGGTAGAGGCGGCTGGTGCTA	85	WNV_PCR_R11B, <u>CCAACTACCGCAACC</u>ACCTARGAGYACCTGGCTGGTCAT	80–82	
	WNV_PCR_F10A, <u>GCCAACTACCGCAACAC</u>AGAAGAGGGTRCAGGAAGTGAAAGGGTACAC	83–84	WNV_PCR_R12A, <u>CCAACTACCGCAACC</u>AGGTCCCTTCCAKGTYTTCTTYTCCAT	79–82	
	WNV_PCR_F10B, <u>GCCAACTACCGCAACAC</u>AAAAAAGAGTYCAAGAAGTCAGAGGGTACAC	81	WNV_PCR_R12B, <u>CCAACTACCGCAACC</u>AGGTCCCTTCCAGGTCCTYTTYTCCAT	81–83	
3 (NS5)	WNV_PCR_F13A, <u>GCCAACTACCGCAACAC</u>GCGAACCACCCTACAGGACCTGGAAC	85	WNV_PCR_R15A, <u>CCAACTACCGCAACC</u>AGGYTTTTTCTCYCTCTTTCCCATC	78–80	420–560
	WNV_PCR_F13B, <u>GCCAACTACCGCAACAC</u>GAGAACCACCCATATAGAACCTGGAAC	82	WNV_PCR_R15B, <u>CCAACTACCGCAACC</u>AGGTTTCTTYTCTCTCTTSCCCATC	79–80	
	WNV_PCR_F14A, <u>GCCAACTACCGCAACAC</u>CAATGTGACCACGATGGCCATGAC	82	WNV_PCR_R16A, <u>CCAACTACCGCAACC</u>ACAGAAAGCGAGCTCCKAGCCAC	81–82	
	WNV_PCR_F14B, <u>GCCAACTACCGCAACAC</u>GAATGTYACCACMATGGCCATGAC	80–82	WNV_PCR_R16B, <u>CCAACTACCGCAACC</u>ACAGGAAGCGRGCYCCCAGCCAC	83–85	

[a] The universal tail sequences appended to the 5' ends of all primers are underlined
[b] T_m, melting temperature

the manufacturer's instructions. RNA was eluted in 60 µl and stored at −70°C until used. One negative and one positive extraction control were processed along with each group of 10 samples subjected to RNA extraction.

Selection of target regions for multiplex PCR/LDR. Thirty nine WNV complete genomic sequences available from the GenBank database (accessed in January 2005) were aligned by using the ClustalW algorithm to select optimal primer binding regions. These regions were characterized by a higher degree of conservation among different WNV strains so as to achieve maximum strain coverage. Primer sets with partially overlapping sequences were designed to simultaneously amplify three different regions (Integrated DNA Technology, Coralville, IA). One region was in the coding sequence of nonstructural protein NS2a, and two regions were in nonstructural protein NS5 (respectively, 6, 8, and 8 PCR primers were used, for a total of 22 primers). Each of the three amplicons was ~500 bp. Each primer sequence contained no more than two degenerate positions and had a melting temperature of around 80°C (Table 3).

A major barrier to high-sensitivity multiplexed PCR amplification in other systems has been the exponential increase in potential false amplicons and primer dimers that results from using too many PCR primers in the same reaction mixture. We circumvent this potential pitfall by using PCR primers containing 5' universal tail sequences on the forward (5'-GCCAACTACCGCAACAC-3') and reverse (5'-CCAACTACCGCAACCA-3') primers (Table 3). Primer dimers and short aberrant amplicons do not amplify efficiently because such products form panhandle structures. The correct PCR product is just the right size for efficient amplification (300 to 600 bp), while false longer amplicons do not amplify as well. We have successfully applied this strategy where standard multiplexed PCR has failed (12, 15, 17).

LDR primers were chosen in three different conserved regions within each of the three PCR amplicons and, just as for the PCR primers, designed with the intent of achieving the highest strain coverage. Each LDR primer pair was composed of an upstream probe bearing a 20-mer zip code complement sequence at its 5' end and a downstream probe bearing a 6-carboxyfluorescein fluorophore at its 3' end. The zip code complements are unique 20-mer oligo-

nucleotide sequences complementary to the zip code addresses spotted on the universal DNA microarray (18).

During LDR, both upstream and downstream probes hybridize to the template sequence and ligation occurs only when there is perfect complementarity at the ligation junction (Fig. 1). Forty-nine LDR primers targeting a total of nine regions (three LDR primer pairs per PCR product) were designed, with the aim of detecting as many strains as possible and potential new variants (Table 4).

LDR products ranged from 72 to 88 bp in length. The oligonucleotide probes (Integrated DNA Technology, Coralville, IA) were designed to have similar thermodynamic features and to avoid hairpin and self-dimer formation.

Two-step multiplex RT-PCR. RT reactions were performed in a 60 µl volume with the SuperScript First Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA) and random hexamers. Briefly, 20 µl of RNA extracted from mosquito pools or from clinical samples was mixed with 6 µl of 50 ng/ml random hexamers. 1 µl of a 10 mM deoxynucleoside triphosphate (dNTP) mixture and 1 µl of water and denatured at 65°C for 5 min. After cooling on ice, 1× RT buffer (20 mM Tris-HCl, 50 mM KCl), 5 mM MgCl₂, 0.01 M dithiothreitol, and 2 U of RNaseOUT were added for a 2-min incubation at 25°C. Six units of SuperScript II RT was added, and the mixture was incubated first at 25°C for 10 min and then at 42°C for 50 min. The reaction was terminated by heating at 70°C for 15 min. Degradation of residual RNA or cleavage of RNA-DNA hybrids was achieved by incubation with 6 U of RNase H for 20 min at 37°C.

Five microliters of newly synthesized cDNA was subjected to multiplex PCR amplification in a final volume of 25 µl which contained 1× GeneAmp PCR Gold buffer, 1.5 mM MgCl₂, 200 µM each dNTP, 5 pmol of each PCR primer, and 1 U of Taq polymerase (AmpliTaq Gold; Applied Biosystems, Foster City, CA). After 10 min of incubation at 95°C for Hot Start Taq activation, a total of 40 cycles were performed, each consisting of a denaturation step at 94°C for 30 s, an annealing step at 60°C for 1 min, and a extension step at 72°C for 1 min, with a final extension step at 72°C for 10 min. A nontemplate negative control and a WNV cDNA positive control were included in each round. All PCR thermal cycling was performed in a Perkin-Elmer GeneAmp PCR System 9700 Thermal

TABLE 4. Primers used for LDR

Target region	Downstream primer sequence (5'-3')	T_m (°C)[b]	Allele-specific (upstream) primer[c] sequence (5'-3')	T_m (°C)	LDR product length (bp)
1 (NS2a)	WNV831aCOM, (Phos)CTATCATGCTTGCA CTCCTAGTCCTAGTGTTTGG (6-FAM)	74	WNV831GZp52, (NH₂)TCCGTTCCGCCA GAGGGTGAGTGGACRGCCAAGATC AGCATKCCAG	85–87	80
	WNV831bCOM, (Phos)CTATACTGATTGCT CTGCTAGTYCTGGTGTTTGG(6-FAM)	73–74			
	WNV831cCOM, (Phos)CCATACTGATTGCC CTGCTAGTTCTAGTGTTTGG(6-FAM)	74			
	WNV939aCOM, (Phos)GAGACGTCG,TGCA CTTGGCACTTATGG(6-FAM)	71	WNV939aGZp53, (NH₂)TCGCCGTCCGCT GTCTTTGGCGTTTGCTGAAGCRAAC TCAGGAG	84	73
	WNV939bCOM, (Phos)GAGACGTGGTACA CTTGGCGCTCATGG(6-FAM)	73	WNV939bGZp53, (NH₂)TCGCCGTCCGCT GTCTTTGGCTTTCGCAGAATCYAAT TCRGGAG	82–84	
	WNV939cCOM, (Phos)GAGATGTGGTGCA TCTGGCGCTCATGG(6-FAM)	73	WNV939cGZp53, (NH₂)TCGCCGTCCGCT GTCTTTGGCCTTTGCAGAATCMAAC TCAGGAG	83–84	
	WNV1021aCOM, (Phos)GACCAACCAAGA GAGTATTTTGCTCATGCTTG(6-FAM)	71	WNV1021aGZp54, (NH₂)GACCAGGCCTC GACCCACCCGGTGGCTTCCTTTTTG AAGGCAAGGTG	87	79
	WNV1021bCOM, (Phos)GACCAACCAGGA GAACATYTTGTTGATGTTGG(6-FAM)	71–73	WNV1021bGZp54, (NH₂)GACCAGGCCTC GACCCACCCGGTGGCATCGTTTCTC AAAGCGAGATG	87	
	WNV1021cCOM, (Phos)GACCAACCAAGA AAACATTCTGCTGATGTTGG(6-FAM)	71	WNV1021cGZp54, (NH₂)GACCAGGCCTC GACCCACCCGGTAGCATCATTTCTC AAGGCGAGATG	86	
2 (NS5)	WNV5341aCOM, (Phos)GTTGGAATATTGT TACCATGAAGAGTGGAGTCGACGTC (6-FAM)	76	WNV5341aGZp55, (NH₂)GCCACGCTGCC AGGACTCCGATGAAGAACCACAACT GGTGCAGAGCTATG	86	88
	WNV5341bCOM, (Phos)GATGGAACATTG TCACCATGAAGAGYGGRGTGGATGT G(6-FAM)	77–79	WNV5341bGZp55, (NH₂)GCCACGCTGCC AGGACTCCGATGAAGAGCCCCAACT RGTGCAAAGTTATG	85–86	
			WNV5341cGZp55, (NH₂)GCCACGCTGCC AGGACTCCGATGAAGAGCCCCAGCT AGTGCAGAGCTATG	88	
	WNV5427aCOM, (Phos)AGTCATCGTCAAG TGCCGAGGTAGAAGAACACCGC(6- FAM)	78	WNV5427aGZp56, (NH₂)TCCGGTCTTGG TCGCTTCGCGCGAGCGACACACTGC TCTGTGACATTGGAG	88	86
	WNV5427bCOM, (Phos)AGTCYTCRTCAAG TGCTGAGGTTGAAGAGCATAGG(6- FAM)	74–76	WNV5427bGZp56, (NH₂)TCCGGTCTTGG TCGCTTCGCTGYTGCGAYACYCTCC TTTGTGACATCGGAG	85–88	
	WNV5427cCOM, (Phos)AGTCCTCATCAAG TGCTGAAGTTGAAGAACATAGA(6- FAM)	72	WNV5427cGZp56, (NH₂)TCCGGTCTTGG TCGCTTCGCTGTTGTGACACTCTCCT TTGTGATATCGGAG	85	
	WNV5548aCOM, (Phos)CCAAAGTGATTGA GAAGACGGAAACACTCC(6-FAM)	69	WNV5548aCZp57, (NH₂)GTCTTCGCGGT GGGTGCCTGCTGCATCAAAGTGCTA TGCCCTTACATGC	86	79
	WNV5548bCOM, (Phos)CRAAAGTCATAG AGAAGATGGAGCTRCTCC (6-FAM)	69–72	WNV5548bCZp57, (NH₂)GTCTTCGCGGT GGGTGCCTGTTGYGTGAAGGTRCTS TGCCCCTACATGC	86–88	
	WNV5548cCOM, (Phos)CAAAGGTCATAG AAAAGATGGAGCTGCTCC(6-FAM)	71			
3 (NS5)	WNV6094aCOM, (Phos)GAGTCAAATACGT CCTCAATGAGACCACGAACTGGCTG (6-FAM)	78	WNV6094aGZp58, (NH₂)GCGTTTGGTTG GCTGCGGACCACAAAAGGCTCCAGA GCCTCCAGAAG	86	
	WNV6094hCOM, (Phos)GAGTGAAGTATG TGCTCAAYGAAACCACCAAYTGGTTG (6-FAM)	75–77	WNV6094bGZp58, (NH₂)GCGTTTGGTTG GCTGCGGACCACGAAAGCTCCBGAA CCGCCAGAAG	86–87	84
	WNV6094cCOM, (Phos)GAGTGAAGTACG TGCTCAAYGAGACCACCAAYTGGTTG (6-FAM)	77–79			
	WNV6094dCOM, (Phos)GAGTTAAGTATGT GCTCAATGAAACCACCAATTGGTTG (6-FAM)	73			

TABLE 4—Continued

Target region	Downstream primer, sequence (5'-3')	T_m (°C)[b]	Allele-specific (upstream) primer,[a] sequence (5'-3')	T_m (°C)	LDR product length (bp)
	WNV6168aCOM, (Phos)TGYTCCCGGGAG GAATTTATYGGAAAAGTCAACAG(6-FAM)	73–76	WNV6168aGZp59, (NH₂)GGACCTCGGCC ACGCTCTGCTTAGCCCGCGAYAARA AACCCAGGATG	86–88	82
	WNV6168bCOM, (Phos)TGCTCGCGAGAG GAATTTATAAGGAAGGTCAATAG(6-FAM)	73	WNV6168bGZp59, (NH₂)GGACCTCGGCC ACGCTCTGCCTGGCACGAGAAAAG CGTCCCAGAATG	88	
	WNV6168cCOM, (Phos)TGCTCTCGAGAGG ARTTCATAAGAAAGGTCAACAG(6-FAM)	73–74	WNV6168cGZp59, (NH₂)GGACCTCGGCC ACGCTCTGCCTGGCCAGAGAAAAC GTCCCAGAATG	87	
	WNV6244aCOM, (Phos)GGAAGAACGCCC GGGAAGCTGTAGAGGATCC(6-FAM)	77	WNV6244aTZp60, (NH₂)GCCACTCGTCC GTCCGCCACAGGAGCGATGTTTGAA GAACAGAACCAAT	85	80
	WNV6244bCOM, (Phos)GGAGGAGTGCTA GAGAAGCGGTTGAAGATCC(6-FAM)	75	WNV6244bTZp60, (NH₂)GCCACTCGTCC GTCCGCCACGGGTGCCATGTTTGAA GARCAGAAYCAAT	85–86	
	WNV6244cCOM, (Phos)GGAGGAGCGCCA GAGAAGCAGTTGAAGATCC(6-FAM)	76			

[a] The zip code complements on the upstream primers are in bold.

[b] T_m, melting temperature

Cycler (Applied Biosystems, Foster City, CA). The amplification products were visualized by electrophoresis in a 2% agarose gel.

One-step multiplex RT-PCR. Alternatively, the RNA extracted from clinical samples was subjected to a one-step multiplex RT-PCR (OneStep RT-PCR kit; Qiagen, Valencia, CA). Briefly, 15 µl of RNA was added to a 50-µl final volume containing 1× OneStep RT-PCR buffer, 0.4 mM dNTPs, 0.6 µM each PCR primer, and 2 µl of OneStep RT-PCR Enzyme Mix. RT (at 50°C for 30 min) was followed by a denaturation step at 95°C for 15 min and 45 cycles of amplification (94°C for 30 s, 60°C for 1 min, and 72°C for 1 min) with a final extension step at 72°C for 10 min.

Multiplex LDR. LDRs were carried out in a final volume of 20 µl containing 5 µl of amplified DNA, 1× LDR buffer (20 mM Tris [pH 7.6], 10 mM MgCl₂, 100 mM KCl, 1 mM NAD, 1 mM dithiothreitol, 250 fmol of each LDR primer, and 0.01 µM AK16D DNA ligase (4, 25).

LDR mixtures were thermally cycled for 20 cycles of 30 s at 94°C and 4 min at 64°C. Prior to LDRs, a mixture containing 7.5 pmol of each LDR primer was phosphorylated in a 30-µl kinase reaction mixture containing 1× T4 ligase buffer (50 mM Tris-HCl, 10 mM MgCl₂, 10 mM dithiothreitol, 1 mM ATP, 25 µg/ml bovine serum albumin) and 10 U of T4 kinase (New England BioLabs, Ipswich, MA). The mixture was incubated at 37°C for 60 min, followed by 10 min of incubation at 65°C and storage at 4°C.

Capillary electrophoresis (CE). After the LDR, the mixture was diluted 1:10 in water and 2 µl was added to 8 µl of a CE master mixture containing Hi-Di Formamide (Applied Biosystems, Foster City, CA) and 0.3 µl of GeneScan 500 LIZ size standard (Applied Biosystem, Foster City, CA). The CE mixture was denatured at 94°C for 2 min and chilled on ice. The LDR products were analyzed on a 3730 DNA analyzer (Applied Biosystems, Foster City, CA). A sample was considered WNV positive when a minimum of two LDR products in any of the three PCR amplicons was detected by CE.

Universal DNA microarray spotting and hybridization conditions. Universal microarrays were prepared by spotting unique 20-mer zip code oligonucleotides (12, 14, 15, 18) on activated Codelink slides (GE Healthcare, Piscataway, NJ) with a QArrayMini robotic array printer (Genetix, Boston, MA) according to the manufacturer's instructions. The zip code addresses were plated into a 384-well microplate in 50 mM sodium phosphate (pH 8.5) at a final concentration of 25 µM. A 1 µM concentration of a fiducial oligonucleotide was added to the printing mixture in each well and cospotted with each zip code address. A carboxy-X-rhodamine-labeled fiducial complement was included in the hybridization mixture to determine the position and quality of each spot. Robotic printing was carried out at 10°C and 50 to 60% humidity. Printed slides were incubated in a saturated NaCl chamber overnight and then treated with a blocking solution (0.1 M Tris, 50 mM ethanolamine [pH 9.0]) to block residual reactive carboxyl groups. The slides were washed with 4× SSC (20× SSC is 3 M sodium chloride and 0.3 M sodium citrate, pH 7.0)–0.1% sodium dodecyl sulfate (SDS) and spin dried. Each printing layout contained a total of 96 zip code addresses spotted in duplicate (see Fig. 4D). Nine zip code addresses were

designed to hybridize the zip code complements appended to the WNV-specific upstream LDR primers. The other zip codes on the array were complementary to zip code complements on LDR primers specific for other blood-borne viral pathogens (dengue virus and other hemorrhagic fever viruses). LDR products from a subset of nine mosquito pool samples and from positive and negative controls were denatured at 94°C for 3 min and chilled on ice prior to hybridization to the arrays. The hybridization solution consisted of the entire volume of the LDR products, 5× SSC, 0.1% SDS, 0.1 mg/ml salmon sperm DNA (Fisher Scientific), and a 5 nM concentration of the fiducial complement in a total volume of 30 µl. The hybridization solution was applied to the microarray slide with a multichamber ProPlate Slide Module (Grace Bio-Labs, Bend, OR). The hybridizations were carried out in the dark on a rocking platform within a hybridization oven (Lab-line; VWR West Chester, PA) at 60°C for 2 h. The slides were rinsed with 5× SSC and washed with 1× SSC–0.1% SDS at 60°C for 15 min. Two more washes followed, first with 0.2× SSC at 22°C for 1 min and then with 0.1× SSC at 22°C for 1 min. The slides were spun dry and scanned on a ProScanArray microarray scanner (Perkin-Elmer, Wellesley, MA).

LOD. The limit of detection (LOD) of the assay was measured with a WNV load panel containing plasma samples spiked with defined dilutions of NY99 virus. The panel titers ranged from 180,000 to 0.15 PFU/ml (quantified by standard plaque assay). Dilutions ranging from 1,800 PFU/ml to 0.07 PFU/ml were tested. For each dilution, RNA was extracted from a 140-µl aliquot with the QIAamp viral RNA mini kit (Qiagen, Valencia, CA) as described above. The RNA was subjected to either the one-step or the two-step multiplex RT-PCR as described above, followed by multiplexed LDR and CE as described earlier. Thus, a dilution of 1,800 PFU/ml corresponds to a 1 OD of 63 PFU for the one-step method (a 140-µl aliquot was extracted into a final volume of 60 µl of RNA, of which 15 µl was used for the RT-PCR) and 28 PFU was used for the two-step method (a 140-µl aliquot was extracted into a final volume of 60 µl of RNA, of which 20 µl was used in the RT step in a total volume of 60 µl. Of this 60-µl volume of cDNA, 2 µl was used in the PCR step).

RESULTS

PCR/LDR/CE primer selection and validation. A multiplex PCR/LDR/CE assay was developed for detecting WNV based on the simultaneous screening of three WNV genomic regions (Fig. 1). The three target regions were initially tested separately by performing PCR/LDR/CE assays with primers specific for each region (Fig. 2A to C). In this way, it was possible to evaluate the primer performance for each region and to ensure signal detection from each of the expected total of nine LDR products. In the initial evaluation phase, performed with
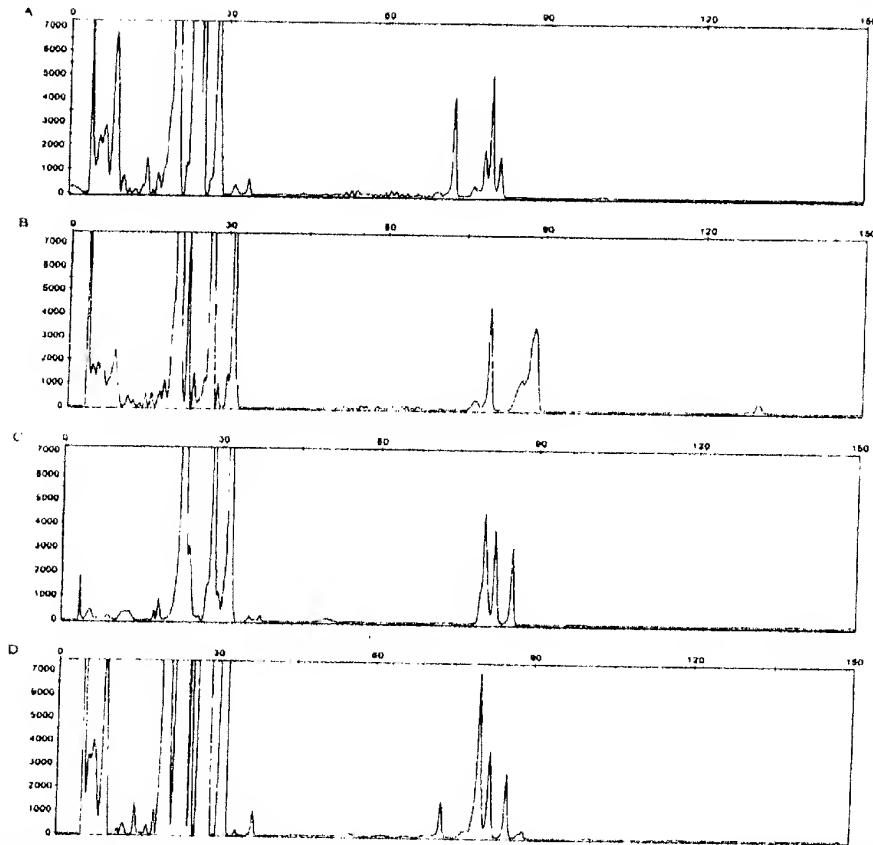
FIG 2 Representative CE profiles of ligation products from the PCR/LDR (A to C) CE profiles from a uniplex PCR/LDR for each of the three regions of WNV (A) Peaks at 72, 79, and 80 bases represent LDR products from regions WNV9J9, WNV1021, and WNV831, respectively. (B) Peaks at 79, 86, and 88 bases represent LDR products from regions WNV5548, WNV5427, and WNV5341, respectively (C) Peaks at 80, 82, and 84 bases represent LDR products from regions WNV6244, WNV6168, and WNV6094, respectively (D) CE cumulative profile obtained after a multiplex PCR/LDR when all three regions are amplified and detected together Note that the peaks at 79 and 80 bases merge together to give a single peak at 80 bases and that the peak at 86 bases merges with the peak at 88 bases, giving a small peak at around 86 bases Fluorescence intensity is indicated on the *y* axis, and the number of bases is indicated on the *x* axis.

WNV cultures, as well as WNV-positive plasma samples obtained from the American Red Cross, the primers which failed to produce either the PCR amplicon or one of the LDR products were discarded and replaced with newly designed primers (data not shown).

Region-specific LDR products for each region produced peaks at 72, 79, and 80 bases for region 1 (along with a minor peak at 81 bases arising from ligation of a primer with a single base difference at one of the ligation sites); 79, 86, and 88 bases for region 2, and 80, 82, and 84 bases for region 3 (Fig. 2A to C) Figure 2D shows the CE profile obtained when the LDR was multiplexed for all three WNV regions. LDR products with identical lengths but different sequences (due to the use of

degenerate oligonucleotides) can migrate at separate positions on CE, resulting in broadened peaks that may overlap nearby peaks. The algorithm for the identification of a positive sample requires detecting the presence of at least two LDR products from any one amplicon or one LDR product from any two amplicons, i.e., at least two separate peaks. No specific peak need be present. The cumulative profile from multiplexed LDR was sufficient to positively identify a sample

LOD. To evaluate the LOD of the WNV multiplex PCR/LDR/CE assay, we tested a viral load panel containing plasma samples spiked with serial dilutions of WNV. The LOD was determined after performing RNA extraction, RT, and multiplex PCR/LDR/CE, and it was calculated both for the assay
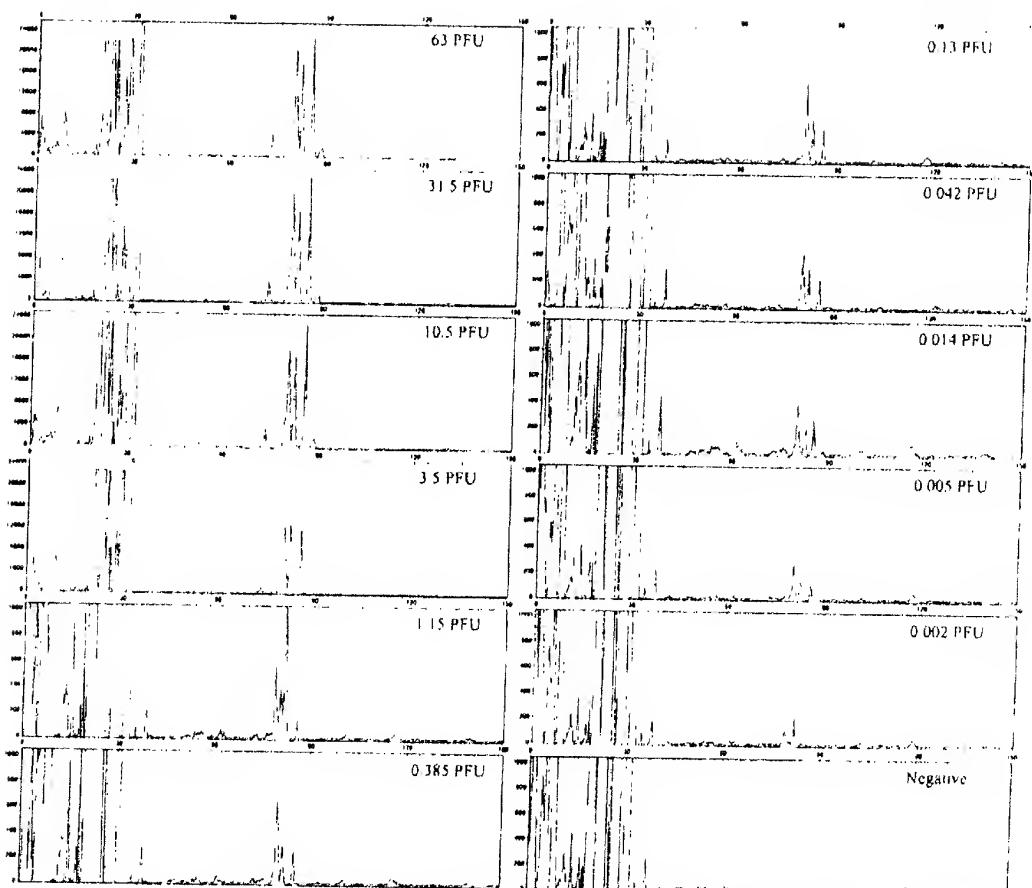
FIG 3 Determination of the sensitivity of the one-step multiplex PCR/LDR/CE assay CE profiles were obtained after RNA extraction and one-step multiplex RT-PCR/LDR of the WNV load panel CE signals below a threshold of 200 fluorescence units were considered negative. Fluorescence intensity is indicated on the y axis, and the number of bases is indicated on the x axis.

that used the two-step approach and for the one-step multiplex RT-PCR

In the two-step approach, the RT was performed with random hexamers, followed by PCR amplification with region-specific primers This approach allowed the detection of the sample dilutions from the viral load panel containing 11 PFU/ml. This corresponds to a LOD of 0.017 PFU. Previous studies using preparations of the NY99 virus strain grown in Vero cells have consistently shown that 1 PFU represents 500 copies (30; R. Lanciotti personal communication). Therefore, based upon these calculations, PCR/LDR has an LOD of approximately eight genome copies.

The one-step multiplex RT-PCR approach allowed the detection of the viral load panel dilutions containing 0 15 PFU/

ml, about 70 times less concentrated than that detected by the two-step method. The LOD with this method corresponds to 0.005 PFU or 2.5 genome copies (Fig. 3). This compares favorably to other detection systems, including those that detect both lineages 1 and 2, such as the FDA-licensed PROCLEIX System (21, 23, 24).

Sensitivity and specificity of multiplex PCR/LDR/CE. The sensitivity of the multiplex PCR/LDR/CE system was determined with WNV cultures and environmental and clinical samples. A sample was considered WNV positive when a minimum of two LDR products in any of the three PCR amplicons was detected. WNV cultures included 34 strains from 19 countries (Table 1) which belonged to both lineages 1 and 2, as well as the Kunjin and Rabensburg viruses. All of the strains tested

positive, except for the Rabensburg virus and two Indian isolates. Both the Rabensburg virus and the two Indian isolates nevertheless produced positive PCR amplification products visible after gel electrophoresis, indicating that the LDR did not work.

To evaluate the specificity of the method, seven other flaviviruses were tested, as listed in Table 2. Although four of them (St. Louis encephalitis virus, yellow fever virus, Murray valley fever virus, and JEV) produced PCR amplification products detected by agarose gel electrophoresis, no ligation products were obtained after the LDR. Ninety-eight pooled mosquito homogenates which had previously tested positive according to the NYC DOHMH were subjected to the two-step multiplex PCR/LDR/CE assay. All but one sample produced a positive signal, for a sensitivity of about 99%. Twenty WNV-negative mosquito pools were also tested, and no false positives were found.

Fifty WNV-positive plasma samples with a representative range of concentrations (a minimum of 100 copies/ml) were obtained from the Gulf Coast Regional Blood Center in Texas. These samples were subjected to RNA extraction and were tested in parallel by both the two-step and one-step multiplex PCR/LDR/CE methods. While the one-step approach detected WNV RNA in all 50 samples with 100% sensitivity, the two-step protocol displayed 82% sensitivity (41 out of 50 positive samples detected). Ninety-two additional WNV-negative plasma samples, together with another 20 dengue virus-positive but WNV-negative samples (obtained from CDC, Puerto Rico), were included in the analysis. No false positives were detected from any of the total of 112 WNV-negative plasma samples, providing 100% specificity.

Universal DNA microarray. A subset of nine WNV-positive mosquito pool samples was tested with the universal DNA microarray as an alternative readout system. Successful ligation of the LDR primers results in the formation of LDR products that bear a zip code complement at the 5' end and a fluorescent label at the 3' end. The universal DNA microarray permits the detection of the ligation products via hybridization of the zip code complements to zip codes spotted on the array. The results obtained with these samples showed that the universal array could detect a fluorescent signal from each of the nine different LDR products which correctly hybridized to their designated addresses on the array (Fig. 4). This indicates that the assay can be performed by using either CE or a universal array as the final readout.

DISCUSSION

In this report, we describe the development of a new WNV detection method based on multiplex RT-PCR followed by LDR. Our detection strategy was based on finding regions in the WNV genome that were most invariant among the different strains belonging to both lineages. We then designed PCR primers that had the required specificity to amplify WNV-specific RNA (after RT) while tolerating sequence variation without amplifying the far more abundant human RNA. Likewise, the LDR primers were designed to specifically ligate, even if the target sequence varied in up to three positions. The high sensitivity of the initial RT-PCR step, with degenerate primers, allows some tolerance to mismatches, which is complemented by the high specificity of the
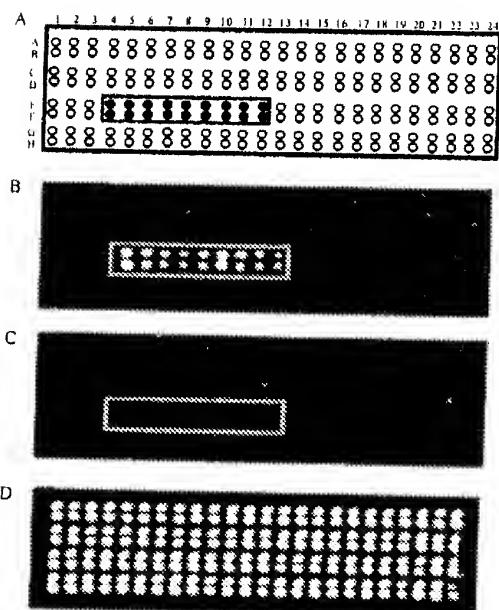

FIG. 4 Representative WNV detection with a universal DNA microarray. (A) Schematic of the microarray. The green box includes the area where the nine WNV-specific zip codes were spotted in duplicate (E4 to E12 and F4 to F12), with red, green, and blue spots indicating the three sets of zip codes corresponding to LDR products for amplicon regions 1, 2, and 3, respectively. Other zip codes on the array are designated for detecting dengue virus (yellow spots) and other hemorrhagic fever viruses (white spots). (B) LDR products from a representative WNV-positive mosquito pool sample revealing hybridization of LDR products from each region to the correct zip codes. The colors represent the fluorescence intensity at each spot, white being the strongest and blue being the weakest. (C) LDR products from a WNV-negative control mosquito pool sample. (D) Hybridization of a carboxy-X-rhodamine-labeled fiducial complement internal control to verify uniform spotting of zip codes.

LDR step. LDR uses an exquisitely specific thermostable AK16D DNA ligase that permits ligation only when the sequence at the junction between the paired oligonucleotides is complementary to the template sequence. This type of assay is ideal for multiplexing, since several primer sets can ligate along a template without the interference encountered in polymerase-based assays (18, 19).

The multiplex RT-PCR/LDR/CE test was evaluated with both mosquito pools and clinical samples, with the clinical samples being subjected to both one-step and two-step RT-PCR protocols. The sensitivity obtained with the mosquito pool samples was 98%; the only sample which gave a negative result was retested at the NYC DOHMH, where it was confirmed as negative, suggesting possible sample degradation.

The WNV-positive clinical samples tested with the one-step protocol, which uses target-specific primers for RT and PCRs, were detected with a sensitivity of 100%. On the other hand,

with the two-step method, where random hexamers are used for the RT step, 82% of the samples gave a positive result. The higher sensitivity of the one-step versus the two-step approach may result from the better performance of the former over the latter method, which was demonstrated during LOD testing (Fig. 3). Although the sensitivity obtained by the two-step approach with mosquito pools did not seem to be affected by this drawback, it has been reported that for low mRNA levels (like those expected to be found in clinical samples versus mosquito pools), gene-specific priming provides a more sensitive method (35).

When the test was evaluated on 34 different WNV strains, the LDR failed to detect the Rabensburg strain and the two Indian isolates belonging to clade 1c or, as new evidence shows, forming a distinct fifth lineage (9). Genomic sequences for these isolates were not available when the LDR primers were designed. Alignment of the Rabensburg and IND 804994 strain sequences, which have since been published, reveal that they are too divergent to successfully anneal with the initial primers. Since LDRs can be highly multiplexed without compromising the ligation efficiency (14, 29), primers permitting the detection of these isolates may easily be incorporated into future versions of the assay. The flexibility of the technique will permit the expansion of the assay to include emerging new WNV strains in a similar manner.

Over the past few years, the PCR/LDR approach has been used for several applications in our laboratories (12, 14–16, 19). The use of LDR primers with specific sequences appended, termed "zip code" complements, has enabled the detection of LDR products through a universal DNA microarray containing designated addressable zip codes (16, 18). A universal DNA microarray offers the advantage of being completely programmable and permits the inclusion of new genomic target sequences without redesigning the array. In addition, different pathogens can be detected simultaneously since the hybridization event is mediated by the spotted zip code and zip code complements on the LDR primers in place of the actual pathogen's genomic sequence. By uncoupling pathogen detection from pathogen identification, the same type of array can be used simultaneously for different organisms without changing the spotted probes.

Our group recently demonstrated the utility of PCR/LDR/CE in the multiplexed detection of blood-borne bacterial infectious agents (29). Due to the frequently nonspecific clinical symptoms of viral infections and the overlap of different arboviruses in the same geographic area, we envision a similar approach to the detection of blood-borne viral pathogens, both in clinical specimens and for environmental surveillance. The use of a multiplex RT-PCR/LDR for detection and a universal DNA microarray for identification represents a convenient tool given the frequent sequence variation in RNA viruses which may necessitate additions to the detection primers used. This approach can assist epidemiologists in rapidly tracking unknown and emerging strains of HFV. We have designed the same type of test for the detection and serotype determination of dengue virus in clinical samples (unpublished data), as well as other hemorrhagic fever viruses, paving the way for a comprehensive viral detection method.

REFERENCES

1. Bakonyi, T., Z. Hubalek, I. Rudolf, and N. Nowotny. 2005. Novel flavivirus or new lineage of West Nile virus, central Europe. Emerg. Infect. Dis. 11:225–231.
2. Bakonyi, T., E. Ivanics, K. Erdelyi, K. Ursu, E. Ferenczi, H. Weissenbock, and N. Nowotny. 2006. Lineage 1 and 2 strains of encephalitic West Nile virus, central Europe. Emerg. Infect. Dis. 12:618–623.
3. Barany, F. 1991. Genetic disease detection and DNA amplification using cloned thermostable ligase. Proc. Natl. Acad. Sci. USA 88:189–193.
4. Barany, F., and D. H. Gelfand. 1991. Cloning, overexpression and nucleotide sequence of a thermostable DNA ligase-encoding gene. Gene 109:1–11.
5. Beasley, D. W., C. T. Davis, H. Guzman, D. L. Vanlandingham, A. P. Travassos da Rosa, R. E. Parsons, S. Higgs, R. B. Tesh, and A. D. Barrett. 2003. Limited evolution of West Nile virus has occurred during its southwesterly spread in the United States. Virology 309:190–195.
6. Beasley, D. W., C. T. Davis, M. Whiteman, B. Granwehr, R. M. Kinney, and A. D. Barrett. 2004. Molecular determinants of virulence of West Nile virus in North America. Arch. Virol. Suppl. 18:35–41.
7. Beasley, D. W., L. Li, M. T. Suderman, and A. D. Barrett. 2001. West Nile virus strains differ in mouse neurovirulence and binding to mouse or human brain membrane receptor preparations. Ann. N. Y. Acad. Sci. 951:332–345.
8. Biggerstaff, B. J., and L. R. Petersen. 2003. Estimated risk of transmission of the West Nile virus through blood transfusion in the US, 2002. Transfusion 43:1007–1017.
9. Bondre, V. P., R. S. Jadi, A. C. Mishra, P. N. Yergolkar, and V. A. Arankalle. 2007. West Nile virus isolates from India: evidence for a distinct genetic lineage. J. Gen. Virol. 88:875–884.
10. Centers for Disease Control and Prevention. 1999. Outbreak of West Nile-like viral encephalitis—New York, 1999. MMWR Morb. Mortal. Wkly. Rep. 48:845–849.
11. Centers for Disease Control and Prevention. 2006. West Nile virus activity—United States, January 1-November 7, 2006. MMWR Morb. Mortal. Wkly. Rep. 55:1204–1205.
12. Cheng, Y.-W., C. Shawber, D. Notterman, P. Paty, and F. Barany. 2006. Multiplexed profiling of candidate genes for CpG island methylation status using a flexible PCR/LDR/Universal Array assay. Genome Res. 16:282–289.
13. Davis, C. T., D. W. Beasley, H. Guzman, R. Raj, M. D'Anton, R. J. Novak, T. R. Unnasch, R. B. Tesh, and A. D. Barrett. 2003. Genetic variation among temporally and geographically distinct West Nile virus isolates, United States, 2001, 2002. Emerg. Infect. Dis. 9:1423–1429.
14. Favis, R., and F. Barany. 2000. Mutation detection in K-ras, BRCA1, BRCA2, and p53 using PCR/LDR and a universal DNA microarray. Ann. N. Y. Acad. Sci. 906:39–43.
15. Favis, R., J. P. Day, N. P. Gerry, C. Phelan, S. Narod, and F. Barany. 2000. Universal DNA array detection of small insertions and deletions in BRCA1 and BRCA2. Nat. Biotechnol. 18:561–564.
16. Favis, R., N. P. Gerry, Y. W. Cheng, and F. Barany. 2005. Applications of the universal DNA microarray in molecular medicine. Methods Mol. Med. 114:25–58.
17. Favis, R., J. Huang, N. P. Gerry, A. Culliford, P. Paty, T. Soussi, and F. Barany. 2004. Harmonized microarray/mutation scanning analysis of TP53 mutations in undissected colorectal tumors. Hum. Mutat. 24:63–75.
18. Gerry, N. P., N. E. Witowski, J. Day, R. P. Hammer, G. Barany, and F. Barany. 1999. Universal DNA microarray method for multiplex detection of low abundance point mutations. J. Mol. Biol. 292:251–262.
19. Khanna, M., P. Park, M. Zirvi, W. Cao, A. Picon, J. Day, P. Paty, and F. Barany. 1999. Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumors. Oncogene 18:27–38.
20. Lanciotti, R. S., G. D. Ebel, V. Deubel, A. J. Kerst, S. Murri, R. Meyer, M. Bowen, N. McKinney, W. E. Morrill, M. B. Crabtree, L. D. Kramer, and J. T. Roehrig. 2002. Complete genome sequences and phylogenetic analysis of West Nile virus strains isolated from the United States, Europe, and the Middle East. Virology 298:96–105.
21. Lanciotti, R. S., A. J. Kerst, R. S. Nasci, M. S. Godsey, C. J. Mitchell, H. M. Savage, N. Komar, N. A. Panella, B. C. Allen, K. E. Volpe, B. S. Davis, and J. T. Roehrig. 2000. Rapid detection of West Nile virus from human clinical specimens, field-collected mosquitoes and avian samples by a TaqMan reverse transcriptase-PCR assay. J. Clin. Microbiol. 38:4066–4071.
22. Lanciotti, R. S., J. T. Roehrig, V. Deubel, J. Smith, M. Parker, K. Steele, B. Crise, K. E. Volpe, M. B. Crabtree, J. H. Scherret, R. A. Hall, J. S. MacKenzie, C. B. Cropp, B. Panigrahy, E. Ostlund, B. Schmitt, M. Malkinson, C. Banet, J.

Weissman, N, Komar, H. M. Savage, W. Stone, T. McNamara, and D. J. Gubler. 1999. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. Science 286:2333–2337.

23. Linke, S., H. Ellerbrok, M. Niedrig, A. Nitsche, and G. Pauli. 2007. Detection of West Nile virus lineages 1 and 2 by real-time PCR. J. Virol. Methods 146:355–358.

24. Linnen, J. M., M. L. Deras, J. Cline, W. Wu, A. S. Broulik, R. E. Cory, J. L. Knight, M. M. Cass, C. S. Collins, and C. Giachetti. 2007. Performance evaluation of the PROCLEIX West Nile virus assay on semi-automated and automated systems. J. Med. Virol. 79:1422–1430.

25. Luo, J., D. E. Bergstrom, and F. Barany. 1996. Improving the fidelity of Thermus thermophilus DNA ligase. Nucleic Acids Res. 24:3071–3078.

26. Mostashari, F., M. L. Bunning, P. T. Kitsutani, D. A. Singer, D. Nash, M. J. Cooper, N. Katz, K. A. Liljebjelke, B. J. Biggerstaff, A. D. Fine, M. C. Layton, S. M. Mullin, A. J. Johnson, D. A. Martin, E. B. Hayes, and G. L. Campbell. 2001. Epidemic West Nile encephalitis, New York, 1999: results of a household-based seroepidemiological survey. Lancet 358:261–264.

27. Niedrig, M., U. Kluckmann, W. Lang, J. Roeder, S, Burk, S. Modrow, and G. Pauli. 1994. Monoclonal antibodies directed against tick-borne encephalitis virus with neutralizing activity in vivo. Acta Virol. 38:141–149.

28. Niedrig, M., S. Linke, H. Zeller, and C. Drosten. 2006. First international proficiency study on West Nile virus molecular detection. Clin. Chem. 52:1851–1854.

29. Pingle, M. R., K. Granger, P. Feinberg, R. Shatsky, B. Sterling, M. Rundell, E. Spitzer, D. Larone, L. Golightly, and F. Barany. 2007. Multiplexed identification of blood-borne bacterial pathogens by use of a novel 16S rRNA gene PCR-ligase detection reaction-capillary electrophoresis assay. J. Clin. Microbiol. 45:1927–1935.

30. Shi, P. Y., E. B. Kauffman, P. Ren, A. Felton, J. H. Tai, A. P. Dupuis II, S. A. Jones, K. A. Ngo, D. C. Nicholas, J. Maffei, G. D. Ebel, K. A. Bernard, and L. D. Kramer. 2001. High-throughput detection of West Nile virus RNA. J. Clin. Microbiol. 39:1264–1271.

31. Shi, P. Y., and L. D. Kramer. 2003. Molecular detection of West Nile virus RNA. Expert Rev. Mol. Diagn. 3:357–366.

32. Smithburn, K. C., T. P. Hughes, A. W. Burke, and J. H. Paul. 1940. A neurotropic virus isolated from the blood of a native of Uganda. Am. J. Trop. Med. s1-20:471–492.

33. Stramer, S. L., C. T. Fang, G. A. Foster, A. G. Wagner, J. P. Brodsky, and R. Y. Dodd. 2005. West Nile virus among blood donors in the United States, 2003 and 2004. N. Engl. J. Med. 353:451–459.

34. Tang, Y., C. Anne Hapip, B. Liu, and C. T. Fang. 2006. Highly sensitive TaqMan RT-PCR assay for detection and quantification of both lineages of West Nile virus RNA. J. Clin. Virol. 36:177–182.

35. Wacker, M. J., and M. P. Godard. 2005. Analysis of one-step and two-step real-time RT-PCR using SuperScript III. J. Biomol. Tech. 16:266–271.

36. White, D. J. 2001. Vector surveillance for West Nile virus. Ann. N. Y. Acad. Sci. 951:74–83.

37. White, D. J., L. D. Kramer, P. B. Backenson, G. Lukacik, G. Johnson, J. A. Oliver, J. J. Howard, R. G. Means, M. Eidson, I. Gotham, V. Kulasekera, and S. Campbell. 2001. Mosquito surveillance and polymerase chain reaction detection of West Nile virus, New York State. Emerg. Infect. Dis. 7:643–649.